PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 9/10, C12Q 1/68, A01K 67/027 // C12N 15/62

A1

(11) International Publication Number: **WO 99/10482**

(43) International Publication Date: 4 March 1999 (04.03.99)

(21) International Application Number: PCT/CA98/00803

(22) International Filing Date: 20 August 1998 (20.08.98)

(30) Priority Data:
2,210,251 25 August 1997 (25.08.97) CA

(71) Applicant *(for all designated States except US)*: McGILL UNIVERSITY [CA/CA]; 845 Sherbrooke Street West, Montréal, Québec H3A 1B1 (CA).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: HEKIMI, Siegfried [CA/CA]; 4235 Powell Avenue, Montréal, Québec H4P 1E5 (CA). LAKOWSKI, Bernard [CA/DE]; Genzentrum, Ludwig–Maximillians Universitaet, Feodor–Lynen–Strasse 25, D–81377 Munich (DE). BARNES, Thomas [AU/US]; 22 Hanson Street #2, Boston, MA 02118 (US). LEMIEUX, Jason [CA/CA]; Apartment 2, 709 Markham Street, Toronto, Ontario M6G 2M2 (CA).

(74) Agents: COTE, France et al.; Swabey Ogilvy Renault, Suite 1600, 1981 McGill College Avenue, Montréal, Québec H3A 2Y3 (CA).

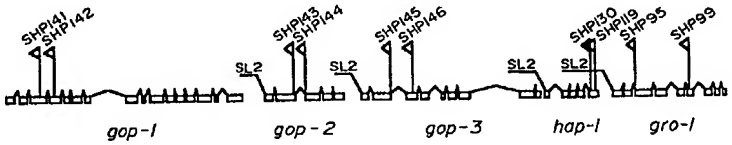
(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: THE C. ELEGANS GRO–1 GENE

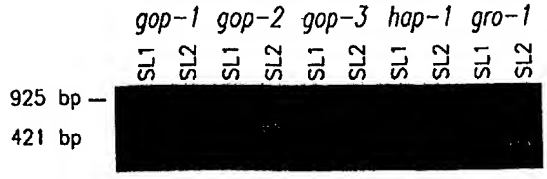
(57) Abstract

The invention relates to the identification of *gro–1* gene and to demonstrate that the *gro–1* gene is involved in the control of a central physiological clock. Also disclosed are four other genes located within the same operon as the *gro–1* gene.



2 kb

A



B

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

THE C. ELEGANS GRO-1 GENE

BACKGROUND OF THE INVENTION

(a) Field of the Invention

The invention relates to the identification of *gro-1* gene and four other genes located within the same operon and to show that the *gro-1* gene is involved in the control of a central physiological clock.

(b) Description of Prior Art

The *gro-1* gene was originally defined by a spontaneous mutation isolated from of a *Caenorhabditis elegans* strain that had recently been established from a wild isolate (J. Hodgkin and T. Doniach, *Genetics* **146**: 149-164 (1997)). We have shown that the activity of the *gro-1* gene controls how fast the worms live and how soon they die. The time taken to progress through embryonic and post-embryonic development, as well as the life span of *gro-1* mutants is increased (Lakowski and Hekimi, *Science* **272**:1010-1013, (1996)). Furthermore, these defects are maternally rescuable: when homozygous mutants (*gro-1/gro-1*) derive from a heterozygous mother (*gro-1/+*), these animals appear to be phenotypically wild-type. The defects are seen only when homozygous mutants derive from a homozygous mother (Lakowski and Hekimi, *Science* **272**:1010-1013, (1996)). In general, the properties of the *gro-1* gene are similar to those of three other genes, *clk-1*, *clk-2* and *clk-3* (Wong *et al.*, *Genetics* **139**: 1247-1259 (1995); Hekimi *et al.*, *Genetics*, **141**: 1351-1367 (1995); Lakowski and Hekimi, *Science* **272**:1010-1013, (1996)), and this combination of phenotypes has been called the Clk ("clock") phenotype. All four of these genes interact to determine developmental rate and longevity in the nematode. Detailed examination of the *clk-1* mutant phenotype has led to the suggestion that there exists a central physiological clock which coordinates

all or many aspects of cellular physiology, from cell division and growth to aging. All four genes have a similar phenotype and thus appear to impinge on this physiological clock.

It would be highly desirable to be provided with the molecular identity of the *gro-1* gene.

SUMMARY OF THE INVENTION

One aim of the present invention is to provide the molecular identity of the *gro-1* gene and four other genes located within the same operon.

In accordance with the present invention there is provided a *gro-1* gene which has a function at the level of cellular physiology involved in developmental rate and longevity, wherein *gro-1* is located within an operon and *gro-1* mutants have a longer life and a altered cellular metabolism relative to the wild-type.

In accordance with a preferred embodiment, the *gro-1* gene of the present invention codes for a GRO-1 protein having the amino acid sequence set forth in Figs. 3A-3B (SEQ ID. NO:2).

The *gro-1* gene is located within an operon which has the nucleotide sequence set forth in SEQ ID NO:1 and which also codes for four other genes, referred as *gop-1*, *gop-2*, *gop-3* and *hap-1* genes.

In accordance with a preferred embodiment, the *gop-1* gene of the present invention codes for a GOP-1 protein having the amino acid sequence set forth in Figs. 13A-13C (SEQ ID. NO:4).

In accordance with a preferred embodiment, the *gop-2* gene of the present invention codes for a GOP-2 protein having the amino acid sequence set forth in Fig. 14 (SEQ ID. NO:5).

In accordance with a preferred embodiment, the *gop-3* gene of the present invention codes for a GOP-3

protein having the amino acid sequence set forth in Figs. 15A-15B (SEQ ID. NO:6).

In accordance with a preferred embodiment, the *hap-1* gene of the present invention codes for a HAP-1 protein having the amino acid sequence set forth in Fig. 16 (SEQ ID. NO:7).

In accordance with a preferred embodiment of the present invention, the *gro-1* gene is of human origin and has the nucleotide sequence set forth in Fig. 8 (SEQ ID. NO:3).

In accordance with a preferred embodiment of the present invention, there is provided a mutant GRO-1 protein which has the amino acid sequence set forth in Fig. 3C.

In accordance with the present invention there is also provided a GRO-1 protein which has a function at the level of cellular physiology involved in developmental rate and longevity, wherein said GRO-1 protein is encoded by the *gro-1* gene identified above.

In accordance with a preferred embodiment of the present invention, there is provided a GRO-1 protein which has the amino acid sequence set forth in Figs. 3A-3B (SEQ ID. NO:2).

In accordance with a preferred embodiment of the present invention, there is provided a GOP-1 protein which has the amino acid sequence set forth in Figs. 13A-13C (SEQ ID. NO:4).

In accordance with a preferred embodiment of the present invention, there is provided a GOP-2 protein which has the amino acid sequence set forth in Fig. 14 (SEQ ID. NO:5).

In accordance with a preferred embodiment of the present invention, there is provided a GOP-3 protein which has the amino acid sequence set forth in Figs. 15A-15B (SEQ ID. NO:6).

In accordance with a preferred embodiment of the present invention, there is provided a HAP-1 protein which has the amino acid sequence set forth in Fig. 16 (SEQ ID. NO:7).

In accordance with the present invention there is also provided a method for the diagnosis and/or prognosis of cancer in a patient, which comprises the steps of:

a) obtaining a tissue sample from said patient;

b) analyzing DNA of the obtained tissue sample of step a) to determine if the human *gro-1* gene is altered, wherein alteration of the human *gro-1* gene is indicative of cancer.

In accordance with the present invention there is also provided a mouse model of aging and cancer, which comprises a gene knock-out of murine gene homologous to *gro-1*.

In accordance with the present invention there is provided the use of compounds interfering with enzymatic activity of GRO-1, GOP-1, GOP-2, GOP-3 or HAP-1 for enhancing longevity of a host.

In accordance with the present invention there is provided the use of compounds interfering with enzymatic activity of GRO-1, GOP-1, GOP-2, GOP-3 or HAP-1 for inhibiting tumorous growth.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1A illustrates the genetic mapping of *gro-1*;

Fig. 1B illustrates the physical map of the *gro-1* region;

Fig. 2A illustrates cosmid clones able to rescue the *gro-1 (e2400)* mutant phenotype;

Fig. 2B illustrates the genes predicted by Genefinder, the relevant restriction sites and the fragments used to subclone the region;

Figs. 3A-3B illustrate the genomic sequence and translation of the *C. elegans gro-1* gene (SEQ. ID. NO:2);

Fig. 3C illustrates the predicted mutant protein;

Fig. 4A illustrates the five genes of the *gro-1* operon (SEQ. ID. NO:1);

Fig. 4B illustrates the transplicing pattern of the five genes of the *gro-1* operon;

Fig. 5 illustrates the alignment of *gro-1* with the published sequences of the *E. coli* (P16384) and yeast (P07884) enzymes;

Fig. 6 illustrates the biosynthetic step catalyzed by DMAPP transferase (MiaAp in *E. coli*, Mod5p in *S. cerevisiae*, and GRO-1 in *C. elegans*);

Fig. 7 illustrates the alignment of the predicted HAP-1 amino acid sequence with homologues from other species;

Fig. 8 illustrates the full mRNA sequence of human homologue of *gro-1* referred to as hgro-1 (SEQ. ID. NO:3);

Fig. 9 illustrates a comparison of the conceptual amino acid sequences for GRO-1 and hgro-1p;

Fig. 10 illustrates a conceptual translation of a partial sequence of the Drosophila homologue of *gro-1* (AA816785);

Fig. 11 illustrates the structure of pMQ8;

Fig. 12 illustrates construction of pMQ18;

Figs. 13A-13C illustrate the genomic sequence and translation of the *gop-1* gene (SEQ. ID. NO:4);

Fig. 14 illustrates the genomic sequence and translation of the *gop-2* gene (SEQ. ID. NO:5);

Figs. 15A-15B illustrate the genomic sequence and translation of the *gop-3* gene (SEQ. ID. NO:6); and

Fig. 16 illustrates the genomic sequence and translation of the *hap-1* gene (SEQ. ID. NO:7).

DETAILED DESCRIPTION OF THE INVENTION

The *gro-1* phenotype

In addition to the previously documented phenotypes, we recently found that *gro-1* mutants were temperature-sensitive for fertility. At 25°C the progeny of these mutants is reduced so much that a viable strain cannot be propagated. In contrast, *gro-1* strains can easily be propagated at 15 and 20°C.

We also discovered that the *gro-1(e2400)* mutation increases the incidence of spontaneous mutations. As *gro-1(e2400)* was originally identified in a non-standard background (Hodgkin and Doniach, *Genetics* **146**: 149-164 (1997)), we first backcrossed the mutations 8 times against N2, the standard wild type strain. We then undertook to examine the *gro-1* strain and N2 for the occurrence of spontaneous mutants which could be identified visually. We focused on the two class of mutants which are detected the most easily by simple visual inspection, uncoordinated mutants (Unc) and dumpy mutants (Dpy). We examined 8200 wild type worms and found no spontaneous visible mutant. By contrast, we found 6 spontaneous mutants among 12500 *gro-1* mutants examined. All mutants produced entirely mutant progeny indicating that they were homozygous.

Sequences of all primers used

Name	Orientation	Sequence (5'-3')	SEQ ID NO:
SHP91	forward	CGAACACTTTATATTTCTCG	SEQ. ID. NO:8
SHP92	reverse	GATAGTTCCCTTCGTTCGGG	SEQ. ID. NO:9
SHP93	forward	TTTCTGGATTTTAACCTTCC	SEQ. ID. NO:10
SHP94	forward	TTTCCGAGAAGTCACGTTGG	SEQ. ID. NO:11
SHP95	reverse	TACAGGAATTTTTGAACGGG	SEQ. ID. NO:12
SHP96	forward	CTTCAGATGACGTGGATTCC	SEQ. ID. NO:13
SHP97	forward	GGAATCCGAAAAAGTGAACT	SEQ. ID. NO:14
SHP98	forward	AAGAGATACACTCAATGGGG	SEQ. ID. NO:15
SHP99	reverse	ATCGATACCACCGTCTCTGG	SEQ. ID. NO:16
SHP109	reverse	TTGAATCTACACTAATCACC	SEQ. ID. NO:17
SHP100	reverse	CCAATTATCTTTTCCAGTCA	SEQ. ID. NO:18
SHP110	forward	ACATTATAAAGTTACTGTCC	SEQ. ID. NO:19
SHP118	forward	TTTTAGTTAAAGCATTGACC	SEQ. ID. NO:20
SHP119	reverse	ACATCTTTATCCATTTCTCC	SEQ. ID. NO:21
SHP120	forward	TGCAAAGGCTCTGGAACTCC	SEQ. ID. NO:22
SHP129	reverse	AAAAACCACTTGATATAAGG	SEQ. ID. NO:23
SHP130	reverse	CATCCAAAAGCAGTATCACC	SEQ. ID. NO:24
SHP134	forward	TTAATTGGATGCAAGCACCCC	SEQ. ID. NO:25
SHP135	reverse	ATTACTATACGAACATTTCC	SEQ. ID. NO:26
SHP138	forward	TTGTAAAGGCGTTAGTTTGG	SEQ. ID. NO:27
SHP139	forward	CAGGAGTATTTGGTGATGCG	SEQ. ID. NO:28
SHP140	forward	CGACGGGGAGAAGGTGACGG	SEQ. ID. NO:29
SHP141	reverse	AAAACTTCTACCAACAATGG	SEQ. ID. NO:30
SHP142	reverse	CGTAATCTCTCTCGATTAGC	SEQ. ID. NO:31
SHP143	reverse	CCGTGGGATGGCTACTTGCC	SEQ. ID. NO:32
SHP144	reverse	TGGATTTGTGGCACGAGCGG	SEQ. ID. NO:33
SHP145	reverse	TTGATTGCCTCTCCTCGTCC	SEQ. ID. NO:34
SHP146	reverse	ATCAACATCTGATTGATTCC	SEQ. ID. NO:35
SHP151	forward	CAGCGAGCGCATGCAACTATATATTG AGCAGG	SEQ. ID. NO:36
SHP159	forward	AATAAATATTTAAATATTCAGATATACC CTGAACTCTACAG	SEQ. ID. NO:37
SHP160	reverse	AAACTGTAGAGTTCAGGGTATATCTG AATATTTAAATATTTATTC	SEQ. ID. NO:38

SHP161	forward	GTACGTGGAGCTCTGCAACTATATATT GAGCAGG	SEQ. ID. NO:39
SHP162	reverse	ATGACACTGCAGGATAGTTCCCTTCG TTCGGG	SEQ. ID. NO:40
SHP163	forward	GTGTTGCATCAGTTCATTCC	SEQ. ID. NO:41
SHP164	forward	GCTGTGCTAGAAGTCAGAGG	SEQ. ID. NO:42
SHP165	reverse	GTTCTCCTTGGAATTCATCC	SEQ. ID. NO:43
SHP170	reverse	AGTATATCTAGATGTGCGAGTCTCTG CCAATT	SEQ. ID. NO:44
SHP171	reverse	AGTAATTGTACATTTAGTGG	SEQ. ID. NO:45
SHP172	forward	ATTAACCTTACTTACTTACC	SEQ. ID. NO:46
SHP173	forward	CTAAACTAAGTAATATAACC	SEQ. ID. NO:47
SHP174	reverse	GTTGATTCTTTGAGCACTGG	SEQ. ID. NO:48
SHP175	forward	AATTCGACCAATTACATTGG	SEQ. ID. NO:49
SHP176	reverse	AACATAGTTGTTGAGGAAGG	SEQ. ID. NO:50
SHP177	forward	AATTAATGGAGATTCTACGG	SEQ. ID. NO:51
SHP178	forward	TCAGCATCTAGAAATGCAGG	SEQ. ID. NO:52
SHP179	reverse	CGAATGTCAACATTCACTGG	SEQ. ID. NO:53
SHP180	forward	CTTAACCTGATGTGTACTCG	SEQ. ID. NO:54
SHP181	forward	ATGAAGCTTTAGAGGATGCC	SEQ. ID. NO:55
SHP182	forward	CGACGAATTTCTGGAGTCGG	SEQ. ID. NO:56
SHP183	reverse	ACTGCATTATCCATTAATCC	SEQ. ID. NO:57
SHP184	reverse	CACCCAAATAACATCTATCC	SEQ. ID. NO:58
SHP185	forward	TTTAACCTCATCTTCGCTGG	SEQ. ID. NO:59
SHP190	forward	ATGTTCCGCAAGCTTGGTTC	SEQ. ID. NO:60
SL1	forward	TTTAATTACCCAAGTTTGAG	SEQ. ID. NO:61
SL2	forward	TTTTAACCCAGTTACTCAAG	SEQ. ID. NO:62

Positional cloning of *gro-1*

gro-1 lies on linkage group III, very close to the gene *clk-1*. To genetically order *gro-1* with respect to *clk-1* on the genetic map, 54 recombinants in the *dpy-17* to *lon-1* interval were selected from among the self progeny of a strain which was *unc-79(e1030) + + clk-1(e2519) lon-1(e678) +/+ dpy-17(e164) gro-1(e2400) + sma-4(e729)*. Three of these showed neither the Gro-1 nor the Clk-1 phenotypes, but carried *unc-79*

and *sma-4*, indicating that these recombination events had occurred between *gro-1* and *clk-1*. From the disposition of the markers, this showed that the gene order was *dpy-17 gro-1 clk-1 lon-1*, and the frequency of events indicated that the *gro-1* to *clk-1* distance was 0.03 map units. In this region of the genome, this corresponds to a physical map distance of ~20 kb.

Several cosmids containing wild-type DNA spanning this region of the genome were tested by microinjection into *gro-1* mutants for their ability to complement the *gro-1(e2400)* mutation (Fig. 1). *gro-1* was mapped between *dpy-17* and *lon-1* on the third chromosome, 0.03 m.u. to the left of *clk-1* (Fig. 1A).

Based on the above genetic mapping, *gro-1* was estimated to be approximately 20 kb to the left of *clk-1*. Eight cosmids (represented by medium bold lines) were selected as candidates for transformation rescue (Fig. 1B). Those which were capable of rescuing the *gro-1(e2400)* mutant phenotype are represented as heavy bold lines (Fig. 1B).

Of these, only B0498, C34E10 and ZC395 were able to rescue the mutant phenotype. Transgenic animals were fully rescued for developmental speed. In addition, the transgenic DNA was able to recapitulate the maternal rescue seen with the wild-type gene, that is, mutants not carrying the transgenic DNA but derived from transgenic mothers display a wild type phenotype. The 7 kb region common to the three rescuing cosmids had been completely sequenced, and this sequence was publicly available.

We generated subclones of ZC395 and assayed them for rescue (Fig. 2). The common 6.5 kb region is blown up in part B. B0498 has not been sequenced and therefore its ends can not be positioned and are therefore represented by arrows.

One subclone pMQ2, spanned 3.9 kb and was also able to completely rescue the growth rate defect and recapitulate the maternal effect. The sequences in pMQ2 potentially encodes two genes. However, a second subclone, pMQ3, which contained only the first of the potential genes (named ZC395.7 in Fig. 2A), was unable to rescue.

Furthermore, frameshifts which would disrupt each of the two genes' coding sequences were constructed in pMQ2 and tested for rescue. Disruption of the first gene (in pMQ4) did not eliminate rescuing ability, but disruption of the second gene (in pMQ5) did. This indicates that the *gro-1* rescuing activity is provided by the second predicted gene.

pMQ2 was generated by deleting a 29.9 kb *Spe*I fragment from ZC395, leaving the left-most 3.9 kb region containing the predicted genes ZC395.7 and ZC395.6 (Fig. 2B). pMQ3 was created in the same fashion, by deleting a 31.4 kb *Nde*I fragment from ZC395, leaving only ZC395.7 intact. In pMQ4, a frameshift was induced in ZC395.7 by degrading the 4 bp overhang of the *Apa*I site. A frameshift was also induced in pMQ5 by filling in the 2 bp overhang of the *Nde*I site found in the second exon of ZC395.6. These frameshifts presumably abolish any function of ZC395.7 and ZC395.6 respectively. The dotted lines represent the extent of frameshift that resulted from these alterations.

To establish the splicing pattern of this gene, cDNAs encompassing the 5' and 3' halves of the gene were produced by reverse transcription-PCR and sequenced (Fig. 3).

This revealed that the gene is composed of 9 exons, spans ~2 kb, and produces an mRNA of 1.3 kb. To confirm that this is indeed the *gro-1* gene, genomic DNA was amplified by PCR from a strain containing the *gro-*

1(e2400) mutation and the amplified product was sequenced. A lesion was found in the 5th exon, where a 9 base-pair sequence has been replaced by a 2 base-pair insertion, leading to a frameshift (Fig. 3C). Fig. 3C illustrates those residues which differ from wild type are in bold.

The reading frame continues out-of-frame for another 33 residues before terminating.

Figs. 3A-B illustrate the coding sequence in capital letters, while the introns, and the untranslated and intergenic sequence are in lower case letters. The protein sequence is shown underneath the coding sequence. Position 1 of the nucleotide sequence is the first base after the SL2 *trans*-splice acceptor sequence. Position 1 of the protein sequence is the initiator methionine. All PCR primers used for genomic and cDNA amplification are represented by arrows. For primers extending downstream (arrows pointing right) the primer sequence corresponds exactly to the nucleotides over which the arrow extends. But for primers extending upstream (arrows pointing left) the primer sequence is actually the complement of the sequence under the arrow. In both cases the arrow head is at the 3' end of the primer. The sequence of the two primers which flank *gro-1* (SHP93 and SHP92) are not represented in this figure. Their sequences are: SHP93 TTTCTGGATTTTAACCTTCC (SEQ. ID. NO:10) and SHP92 GATAGTTCCCTTCGTTCGGG (SEQ. ID. NO:9). The wild type splicing pattern was determined by sequencing of the cDNA. Identification of the *e2400* lesion was accomplished by sequencing the *e2400* allele. The *e2400* lesion consists of a 9 bp deletion and a 2 bp insertion at position 1196, resulting in a frameshift.

gro-1 is part of a complex operon (Figs. 3A-3B)

Amplification of the 5' end of *gro-1* from cDNA occurred only when the *trans*-spliced leader SL2 was used as the 5' primer, and not when SL1 was used. SL2 is used for *trans*-splicing to the downstream gene when two genes are organized into an operon (Spieth *et al.*, *Cell* **73**: 521-532 (1993); Zorio et al., *Nature* **372**: 270-272 (1994)). This indicates that at least one gene upstream of *gro-1* is co-transcribed with *gro-1* from a common promoter. We found that sequences from the 5' end of the three next predicted genes upstream of *gro-1* (ZC395.7, C34E10.1, and C34E10.2) all could only be amplified with SL2. Sequences from the fourth predicted upstream gene (C34E10.3), however, could be amplified with neither spliced leader, suggesting that it is not *trans*-spliced. The distance between genes in operons appear to have an upper limit (Spieth *et al.*, *Cell* **73**: 521-532 (1993); Zorio et al., *Nature* **372**: 270-272 (1994)), and no gene is predicted to be close enough upstream of C34E10.3 or downstream of *gro-1* to be co-transcribed with these genes. Our findings suggest therefore that *gro-1* is the last gene in an operon of five co-transcribed genes (Fig. 4).

Nested PCR was used to amplify the 5' end of each gene. SL1 or SL2 specific primers were used in conjunction with a pair of gene-specific primers. cDNA generated by RT-PCR using mixed stage N2 RNA was used as template in the nested PCR. Fig. 4A illustrates a schematic of the *gro-1* operon showing the coding sequences of each gene and the primers (represented by flags) used to establish the *trans*-splicing patterns.

Fig. 4B illustrates the products of the PCR with SL1 and SL2 specific primers for each of the five genes. The sequences of the primers used are as follows: SL1: TTTAATTACCCAAGTTTGAG (SEQ. ID. NO:61), SL2:

TTTTAACCCAGTTACTCAAG (SEQ. ID. NO:62), SHP141:
AAAACTTCTACCAACAATGG (SEQ. ID. NO:30), SHP142:
CGTAATCTCTCTCGATTAGC (SEQ. ID. NO:31), SHP143:
CCGTGGGATGGCTACTTGCC (SEQ. ID. NO:32), SHP144:
TGGATTTGTGGCACGAGCGG (SEQ. ID. NO:33), SHP145:
TTGATTGCCTCTCCTCGTCC (SEQ. ID. NO:34), SHP146:
ATCAACATCTGATTGATTCC (SEQ. ID. NO:35), SHP130:
CATCCAAAAGCAGTATCACC (SEQ. ID. NO:24), SHP119:
ACATCTTTATCCATTTCTCC (SEQ. ID. NO:21), SHP95:
TACAGGAATTTTTGAACGGG (SEQ. ID. NO:12), SHP99:
ATCGATACCACCGTCTCTGG (SEQ. ID. NO:16).

The gene immediately upstream of *gro-1*, has homology to the yeast gene *HAM1*, and we have renamed the gene *hap-1*. We have established its splicing pattern by reverse transcription PCR and sequencing. This revealed that *hap-1* is composed of 5 exons and produces an mRNA of 0.9 kb. We also found that sequences which were predicted to belong to ZC395.7 (now *hap-1*) are in fact spliced to the exons of C34E10.1. This is consistent with our finding that *hap-1* is SL2 spliced as it puts the end of the C34E10.1 very close to the start of *hap-1* (Fig. 4).

The *gro-1* gene product

Conceptual translation of the *gro-1* transcript indicated that it encodes a protein of 430 amino acids highly similar to a strongly conserved cellular enzyme: dimethylallyldiphosphate:tRNA dimethylallyltransferase (DMAPP transferase). Fig. 5 shows an alignment of *gro-1* with the published sequences of the *E. coli* (P16384) and yeast (P07884) enzymes. Residues where the biochemical character of the amino acids is conserved are shown in bold. Identical amino acids are indicated further with a dot. The ATP/GTP binding site and the C2H2 zinc finger site are predicted and not experimental. The point at which the *gro-1(e2400)*

mutation alters the reading frame of the sequence is shown. The two alternative initiator methionines in the yeast sequence, and the putative corresponding methionines in the worm sequence, are underlined.

Database searches also identified a homologous human expressed sequence tag (Genbank ID: Z40724). The human clone has been used to derive a sequence tagged site (STS). This means that the genetic and physical position of the human *gro-1* homologue is known. It maps to chromosome 1, 122.8 cR from the top of Chr 1 linkage group and between the markers D1S255 and D1S2861. This information was found in the UniGene database or the National Center for Biotechnology Information (NCBI). We have sequenced Z40724 by classical methods but found that Z40724 is not a full length cDNA clone as it does not contain an initiator methionine nor the poly A tail. We used the sequence of Z40724 to identify further clones by database searches. We found one clone (Genbank ID: AA332152) which extended the sequence 5′ by 28 nucleotides, as well as one clone (Genbank ID: AA121465) which extended the sequence substantially in the 3′ direction but didn't include the poly A tail. We then used AA121465 to identify an additional clone (AA847885) extending the sequence to the poly A tail. Fig. 8 shows the full sequence with the putative initiator ATG shown in bold and the sequence of Z60724 is shown underlined. A comparison of the conceptual amino acid sequences for GRO-1 and hgro-1p is shown in Fig. 9. Amino acid identities are indicated by a dot. Both sequences contain a region with a zinc finger motif which is shown underlined.

An additional metazoan homologue is represented by Drosophila EST: Genbank accession: AA816785. In *E. coli* and other bacteria, the gene encoding DMAPP trans-

ferase is called *miaA* (a.k.a *trpX*) and is called *mod5* in yeast. DMAPP transferase catalyzes the modification of adenosine 37 of tRNAs whose anticodon begins with U (Fig. 6).

In these organisms the enzyme has been shown to use dimethylallyldiphosphate as a donor to generate dimethylallyl-adenosine (dma^6A37), one base 3' to the anticodon (for review and biochemical characterization of the bacterial enzyme see Persson *et al., Biochimie* **76**: 1152-1160 (1994); Leung et *al., J Biol Chem* **272**: 13073-13083 (1997); Moore and Poulter, *Biochemistry* **36**:604-614 (1997)). In earlier literature this modification is often referred to as isopentenyl adenosine (i^6A37).

The high degree of conservation of the protein sequence between GRO-1 and DMAPP in *S. cerevisiae* and *E. coli* suggest that GRO-1 possesses the same enzymatic activity as the previously characterized genes. The sequence contains a number of conserved structural motifs (Fig. 5), including a region with an ATP/GTP binding motif which is generally referred to as the 'A' consensus sequence (Walker *et al., EMBO J* 1: 945-951 (1982)) or the 'P-loop' (Saraste *et al., Trends Biochem Sci* **15**: 430-434 (1990)).

In addition, at the C-terminal end of the GRO-1 sequence, there is a C2H2 zinc finger motif as defined by the PROSITE database. This type of DNA-binding motif is believed to bind nucleic acids (Klug and Rhodes, *Trends Biochem Sci* **12**: 464-469 (1987)). Although there appears to be some conservation between the worm and yeast sequences in the C-terminus end of the protein (Fig. 5), including in the region encompassing the zinc finger in GRO-1, the zinc finger motif per se is not conserved in yeast but is present in humans (Fig. 9).

In yeast DMAPP transferase is the product of the *MOD5* gene, and exists in two forms: one form which is targeted principally to the mitochondria, and one form which is found in the cytoplasm and nucleus. These two forms differ only by a short N-terminal sequence whose presence or absence is determined by differential translation initiation at two "in frame" ATG codons. (Gillman *et al., Mol & Cell Biol* **11**: 2382-90 (1991)). The *gro-1* open reading frame also contains two ATG codons at comparable positions, with the coding sequence between the two codons constituting a plausible mitochondrial sorting signal (Figs. 3 and 5). It is likely therefore that DMAPP transferase in worms also exists in two forms, mitochondrial and cytoplasmic.

It should be noted, however, that the sequence of hgro-1 shows only one in-frame methionine before the conserved ATP/GTP binding site (Fig. 9). As we cannot be assured to have determined the sequence of the full length transcript, it is possible that further 5' sequence might reveal an additional methionine. Alternatively, in humans, the mechanism by which the enzyme is targeted to several compartments might not involved differential translation initiation. In this context, it should be noted that the sorting signals which can be predicted from the sequence of hgro-1p are predicted to be highly ambiguous by the prediction program PSORT II. Furthermore, a conceptual translation of the Drosophila sequence (AA816785) predicts only one initiator methionine before the ATP/GTP binding site as well as several in-frame stop codons upstream of this start (Fig. 10), suggesting that no additional upstream ATG could serve as translation initiation site. In the figure, stop codons are indicated by <u>stop</u>, methionines are indicated by **Met**, and the conserved ATP/GTP binding site is underlined.

Expression pattern of GRO-1

We have also constructed a reporter gene expressing a fusion protein containing the entire GRO-1 amino acid sequence fused at the C-terminal end to green fluorescent protein (GFP). The promotor of the reporter gene is the sequence upstream of *gop-1* (Figs. 13A-13C), the first gene in the operon (see Fig. 4). The promotor sequence is 306 bp long starting 32 nucleotides upstream of the *gop-1* ATG. It is fused at the exact level upstream of *gro-1* where trans-splicing to SL2 normally occurs.

The genes *gop-2* (Fig. 14) and *gop-3* (Figs. 15A-15B) are also located in the operon (see Fig. 4), the second and third genes in the operon.

We first construct the clone pMQ8 in which *gro-1* is directly under the promoter for the whole operon using the hybrid primers SHP160 (SEQ. ID. NO:38) and SHP159 (SEQ. ID. NO:37) and the flanking primers SHP161 (SEQ. ID. NO:39) and SHP162 (SEQ. ID. NO:40) in sequential reactions each followed by purification of the products and finally cloning into pUC18 (Fig. 11).

Primers SHP151 (SEQ. ID. NO:36) and SHP170 (SEQ. ID. NO:44) where then used to amplify part of the insert in pMQ8 and clone in pPD95.77 (gift from Dr Andrew Fire) which was designed to allow a protein of interest to be transcriptionally fused to Green Fluorescent Protein (GFP) (Fig. 12).

The reporter construct fully rescues the phenotype of a *gro-1(e2400)* mutant upon injection and extrachromosomal array formation, indicating that the fusion to the GFP moiety does not significantly inhibit the function of GRO-1. Fluorescent microscopy indicated that *gro-1* is expressed in most or all somatic cells. Furthermore, the GRO-1::GFP fusion protein is localized

in the mitochondria, in the cytoplasm as well as in the nucleus.

The *hap-1* gene product (Fig. 16)

hap-1 is homologous to the yeast gene *HAM1* as well as to sequences in many organisms including bacteria and mammals (Fig. 7).

The origin of the worm and yeast sequence is as described above and below. The human sequence was inferred from a cDNA sequence assembled from expressed sequence tags (ESTs); the accession numbers of the sequences used were: AA024489, AA024794, AA025334, AA026396, AA026452, AA026502, AA026503, AA026611, AA026723, AA035035, AA035523, AA047591, AA047599, AA056452, AA115232, AA115352, AA129022, AA129023, AA159841, AA160353, AA204926, AA226949, AA227197 and D20115. The *E. coli* sequence is a predicted gene (accession 1723866).

Mutations in *HAM1* increase the sensitivity of yeast to the mutagenic compound 6-N-hydroxylaminopurine (HAP), but do not increase spontaneous mutation frequency (Nostov *et al., Yeast* **12**:17-29 (1996)). HAP is an analog of adenine and *in vitro* experiments suggest that the mechanism of HAP mutagenesis is its conversion to a deoxynucleoside triphosphate which is incorporated ambiguously for dATP and dGTP during DNA replication (Abdul-Masih and Bessman, *J Biol Chem* **261 (5)**: 2020-2026 (1986)). The role of the Haml p gene product in increasing sensitivity to HAP remains unclear.

Explaining the pleiotropy of *miaA* and *gro-1*

Mutations in *miaA*, the bacterial homologue of *gro-1*, show multiple phenotypes and affect cellular growth in complex ways. For example, in *Salmonella typhimurium*, such mutations result in 1) a decreased efficacy of suppression by some suppressor tRNA, 2) a slowing of ribosomal translation, 3) slow growth under

various nutritional conditions, 4) altered regulation of several amino acid biosynthetic operons, 5) sensitivity to chemical oxidants and 6) temperature sensitivity for aerobic growth (Ericson and Björk, *J. Bacteriol.* **166**: 1013-1021 (1986); Blum, *J. Bacteriol.* **170**: 5125-5133 (1988)). Thus, MiaAp appears to be important in the regulation of multiple parallel processes of cellular physiology. Although we have not yet explored the cellular physiology of *gro-1* mutants along the lines which have been pursued in bacteria, the apparently central role of *miaA* is consistent with our findings that *gro-1*, and the other genes with a Clk phenotype, regulate many disparate physiological and metabolic processes in *C. elegans* (Wong et al., *Genetics* **139**: 1247-1259 (1995) ; Lakowski and Hekimi, *Science* **272**: 1010-1013 (1996); Ewbank et *al., Science* **275**: 980-983 (1997)).

In addition to the various phenotypes discussed above, *miaA* mutations increase the frequency of spontaneous mutations (Connolly and Winkler, *J Bacteriol* **173(5)**: 1711-21 (1991); Connolly and Winkler, *J Bacteriol* **171**: 3233-46 (1989)). As described in the previous section we have preliminary evidence that *gro-1(e2400)* also increases the frequency of spontaneous mutations in worms.

How can the alteration in the function of MDAPP transferase result in so many distinct phenotypes? Bacterial geneticists working with *miaA* have generally suggested that this enzyme and the tRNA modification it catalyzes have a regulatory function which is mediated through attenuation (e.g. Ericson and Björk, *J. Bacteriol.* **166**: 1013-1021 (1986)). Attenuation is a phenomenon by which the transcription of a gene is interrupted depending on the rate at which ribosomes can translate the nascent transcript. Ribosomal transla-

tion is slowed in *miaA* mutants, and thus, through an effect on attenuation, could affect the expression of many genes whose expression is regulated by attenuation.

gro-1(e2400) also produces pleiotropic effects and, in addition, displays a maternal-effect, suggesting that it is involved in a regulatory process (Wong *et al.*, *Genetics* **139**: 1247-1259 (1995). However, attenuation involves the co-transcriptional translation of nascent transcripts, which is not possible in eukaryotic cells were transcription and translation are spatially separated by the nuclear membrane. If the basis of the pleiotropy in *miaA* and *gro-1* is the same, then a mechanism distinct from attenuation has to be involved. Below we argue that this mechanism could be the modification by DMAPP transferase of adenine residues in DNA in addition to modification of tRNAs.

A role for *gro-1* in DNA modification?

We observed that *gro-1* can be rescued by a maternal effect, so that adult worms homozygous for the mutation, but issued from mother carrying one wild type copy of the gene display a wild type phenotype, in spite of the fact that such adults are up to 1000 fold larger than the egg produced by their mother. It is unlikely that enough wild type product can be deposited by the mother in the egg to rescue a adult which is 1000 times larger. This observation suggests therefore that *gro-1* can induce an epigenetic state which is not altered by subsequent somatic growth. One of the best documented epigenetic mechanisms is imprinting in mammals (Lalande, *Annu Rev Genet* **30**: 173-196 (1996)) which is believed to rely on the differential methylation of genes (Laird and Jaenisch, *Annu Rev Genet* **30**: 441-464; Klein and Costa, Mutat Res **386**: 103-105 (1997)). Modification of bases in DNA have also been linked to regu-

lation of gene expression in the protozoan *Trypanosoma brucei*. The presence of beta-D-glucosyl-hydroxymethyluracil in the long telomeric repeats of *T. brucei* correlates with the repression of surface antigen gene expression (Gommers-Ampt *et al.*, *Cell* **75**: 112-1136 (1993); van Leeuwen *et al.*, *Nucleic Acids Res* **24**: 2476-2482 (1996)).

gro-1 and *miaA* increase the rate of spontaneous mutations, which is generally suggestive of a role in DNA metabolism, and can be related to the observation that methylation is linked to spontaneous mutagenesis, genome instability, and cancer (Jones and Gonzalgo, *Proc. Natl. Acad. Sci. USA*, **94**: 2103-2105 (1997)).

Does *gro-1* have access to DNA? Studies with *mod5*, the yeast homologue of *gro-1*, have shown that there are two forms of Mod5p, one is localized to the nucleus as well as to the cytoplasm, and the other form is localized to the mitochondria as well as the cytoplasm (Boguta *et al.*, *Mol. Cell. Biol.* **14**: 2298-2306 (1994)). The nuclear localization is striking as isopentenylation of nuclear-encoded tRNA is believed to occur exclusively in the cytoplasm (reviewed in Boguta *et al.*, *Mol. Cell. Biol.* **14**: 2298-2306 (1994)). Furthermore, studies of a gene *maf1* have shown that when *mod5* is mislocalized to the nucleus, the efficiency of certain suppressor tRNA is decreased, an effect known to be linked to the absence of the tRNA modification (Murawski *et al.*, *Acta Biochim. Pol.* **41**: 441-448 (1994)). Finally, as described in the previous section, *gro-1* contains a zinc finger, a nuclei acid binding motif. The zinc finger could bind tRNAs, but as it is in the C-terminal domain of *gro-1* and human hgro-1 that has no equivalent in *miaA*, it is clearly not necessary for the basic enzymatic function. We speculate that it might be necessary to increase the

specificity of DNA binding in the large metazoan genome. It should also be noticed that the second form of Mod5p which is localized to mitochondria also has the opportunity to bind and possibly modify DNA as it has access to the mitochondrial genome. See the previous section entitled "A role for *gro-1* in a central mechanism of physiological coordination" for an alternative possibility as to the function of GRO-1 in the nucleus.

miaA and *gro-1* are found in complex operons

We have found that *gro-1* is part of a complex operon of five genes (Fig. 4). It is believed that genes are regulated coordinately by single promoters when they participate in a common function (Spieth *et al.*, *Cell* **73**: 521-532 (1993)). In some cases, this is well documented. For example, the proteins LIN-15A and LIN-15B which are both required for vulva formation in *C. elegans*, are unrelated products from two genes transcribed in a common operon (Huang et al., *Mol Biol Cell* **5(4)**: 395-411 (1994)). One of the genes in the *gro-1* promoter is *hap-1*, whose yeast homologue has been shown to be involved in the control of mutagenesis (Nostov *et al.*, *Yeast* **12**: 17-29 (1996)). Under the hypothesis that *gro-1* modifies DNA, it suggest an involvement of *hap-1* in this or similar processes. The presence in the same operon also suggest that all five genes might collaborate in a common function. The phenotype of *gro-1* suggests that this function is regulatory. In this context, it should be noted that *miaA* also is part of a particularly complex operon (Tsui and Winkler, Biochimie **76**: 1168-1177 (1994)), although, except for *miaA/gro-1*, there are no other homologous genes in the two operons.

A role for *gro-1* in a central mechanism of physiological coordination

We have speculated that the genes with a Clk phenotype might participate in a central mechanism of physiological coordination, probably including the regulation of energy metabolism. *clk-1* encodes a mitochondrial protein (unpublished observations), and its homologue in yeast has also been shown to be mitochondrial (Jonassen, T (1998) *Journal of Biological Chemistry* **273**:3351-3357). The yeast *clk-1* homologue is involved in the regulation of the biosynthesis of ubiquinone (Marbois, B.N. and Clarke, C.F. (1996) *Journal of Biological Chemistry* **271**:2995-3004). Ubiquinone, also called coenzyme Q, is central to the production of ATP in mitochondria. In worms, however, we have found that *clk-1* is not strictly required for respiration. How might *gro-1* fit into this picture?

One link is that dimethylallyldiphosphate is known to be the precursor of the lipid side-chain of ubiquinone. In bacteria, ubiquinone is the major lipid made from DMAPP. In eukaryotes cholesterol and its derivatives are also made from DMAPP. Interestingly, *C. elegans* requires cholesterol in the growth medium for optimal growth. This link, however, remains tenuous, in particular in the absence of an understanding of the biochemical function of CLK-1.

In several bacteria, the adenosine modification carried out by DMAPP transferase is only the first step in a series of further modification of this base (Persson *et al.*, *Biochimie* **76**: 1152-1160 (1994)). These additional modifications have been proposed to play the role of a sensor for the metabolic state of the cell (Buck and Ames, *Cell* **36**: 523-531 (1984); Persson and Björk, *J. Bacteriol.* **175**: 7776-7785 (1993)). For example, one of the subsequent steps, the synthesis of 2-methylthio-cis-ribozeatin is carried

out by a hydroxylase encoded by the gene *miaE*. When the cells lack *miaE* they become incapable of using intermediates of the citric acid cycle such as fumarate and malate as the sole carbon source.

Another link to energy metabolism springs from the recent biochemical observations of Winkler and co-workers using purified DMAPP transferase (*E. coli* MiaAp) (Leung *et al., J Biol Chem* 272: 13073-13083 (1997)). These investigators observed that the enzyme in competitively inhibited by phosphate nucleotides such as ATP or GTP. Furthermore, using their estimation of K_m of the enzyme and its concentration in the cell, they calculate that the level of inhibition of the enzyme *in vivo*, would exactly allow the enzyme to modify all tRNAs but any further inhibition would leave unmodified tRNAs. This suggests that the exact level of modification of tRNA (or of DNA) could be exquisitely sensitive to the level of phosphate nucleotides. Superficially, this is consistent with the phenotypic observations. The state of mutant cells which lack DMAPP transferase entirely would be equivalent of cells where very high levels of ATP would completely inhibit the enzyme. Such cells might therefore turn down the ATP generating processes in response to the signal provided by undermodified tRNAs (or DNA).

More generally, GRO-1 could act in the crosstalk between nuclear and mitochondrial genomes. The nuclear and mitochondrial genomes both contribute gene products to the mitochondrion energy-producing machinery and these physically separate genomes must therefore exchange information somehow to coordinate their contributions (reviewed in Poyton, R.O. and McEwen J.E. (1996) *Annu. Rev. Biochem.* 65:563-607). Furthermore, the energy producing activity of the mitochondria is essential to the rest of the cell, and the needs of a

particular cell at a particular time must be somehow convey to the organelle to regulate its activity. GRO-1 could participate in this coordination in the following manner. GRO-1 is found in three compartments, the nucleus, the cytoplasm and the mitochondria (see above), and thus has the opportunity to regulate gene expression in more that one way. How could its action coordinate gene expression between compartment? GRO-1 could partition between the mitochondria and the nucleus and its relative distribution could be determined by the amount of RNA (or mtDNA) in the mitochondria (Parikh, V.S. et al. (1987) *Science* **235**:576-580). For example, if the cell is rich in mitochondria, much GRO-1 will be bound there which could result in a relative depletion of activity in the cytoplasm with regulatory consequences on the translation machinery. Binding of GRO-1 in the nucleus could have similar consequences and provide information about nuclear gene expression to the translation machinery.

While the invention has been described in connection with specific embodiments thereof, it will be understood that it is capable of further modifications and this application is intended to cover any variations, uses, or adaptations of the invention following, in general, the principles of the invention and including such departures from the present disclosure as come within known or customary practice within the art to which the invention pertains and as may be applied to the essential features hereinbefore set forth, and as follows in the scope of the appended claims.

WHAT IS CLAIMED IS:

1. A *gro-1* gene which has a function at the level of cellular physiology involved in developmental rate and longevity, wherein *gro-1* is located within an operon and *gro-1* mutants have a longer life and a altered cellular metabolism relative to the wild-type.

2. The *gro-1* gene of claim 1, wherein said operon has the nucleotide sequence set forth in SEQ ID. NO:1.

3. The *gro-1* gene of claim 1, which codes for a GRO-1 protein having the amino acid sequence set forth in Figs. 3A-3B (SEQ ID. NO:2).

4. A *gop-1* gene which codes for a GOP-1 protein having the amino acid sequence set forth in Figs. 13A-13C (SEQ ID. NO:4).

5. A *gop-2* gene which codes for a GOP-2 protein having the amino acid sequence set forth in Fig. 14 (SEQ ID. NO:5).

6. A *gop-3* gene which codes for a GOP-3 protein having the amino acid sequence set forth in Figs. 15A-15B (SEQ ID. NO:6).

7. A *hap-1* gene which codes for a HAP-1 protein having the amino acid sequence set forth in Fig. 16 (SEQ ID. NO:7).

8. The *gro-1* gene of claim 1, wherein said gene is of human origin and which has the nucleotide sequence set forth in Fig. 8 (SEQ ID. NO:3).

9. A GRO-1 protein which has a function at the level of cellular physiology involved in developmental rate and longevity, wherein said GRO-1 protein is encoded by the gene of claim 1 or 2.

10. A mutant GRO-1 protein which has the amino acid sequence set forth in Fig. 3C.

11. A GRO-1 protein which has the amino acid sequence set forth in Figs. 3A-3B (SEQ ID. NO:2).

12. A GOP-1 protein which has the amino acid sequence set forth in Figs. 13A-13C (SEQ ID. NO:4).

13. A GOP-2 protein which has the amino acid sequence set forth in Fig. 14 (SEQ ID. NO:5).

14. A GOP-3 protein which has the amino acid sequence set forth in Figs. 15A-15B (SEQ ID. NO:6).

15. A HAP-1 protein which has the amino acid sequence set forth in Fig. 16 (SEQ ID. NO:7).

16. A method for the diagnosis and/or prognosis of cancer in a patient, which comprises the steps of:
a) obtaining a tissue sample from said patient;
b) analyzing DNA of the obtained tissue sample of step a) to determine if the human *gro-1* gene is altered, wherein alteration of the human *gro-1* gene is indicative of cancer.

17. A mouse model of aging and cancer, which comprises a gene knock-out of murine gene homologous to *gro-1* gene of claims 1 to 3.

18. The use of compounds interfering with enzymatic activity of GRO-1 of claim 9, 10 or 11 for enhancing longevity of a host.

19. The use of compounds interfering with enzymatic activity of GOP-1 of claim 12 for enhancing longevity of a host.

20. The use of compounds interfering with enzymatic activity of GOP-2 of claim 13 for enhancing longevity of a host.

21. The use of compounds interfering with enzymatic activity of GOP-3 of claim 14 for enhancing longevity of a host.

22. The use of compounds interfering with enzymatic activity of HAP-1 of claim 15 for enhancing longevity of a host.

23. The use of compounds interfering with enzymatic activity of GRO-1 of claim 9, 10 or 11 for inhibiting tumorous growth.

24. The use of compounds interfering with enzymatic activity of GOP-1 of claim 12 for inhibiting tumorous growth.

25. The use of compounds interfering with enzymatic activity of GOP-2 of claim 13 for inhibiting tumorous growth.

26. The use of compounds interfering with enzymatic activity of GOP-3 of claim 14 for inhibiting tumorous

growth.

27. The use of compounds interfering with enzymatic activity of HAP-1 of claim 15 for inhibiting tumorous growth.

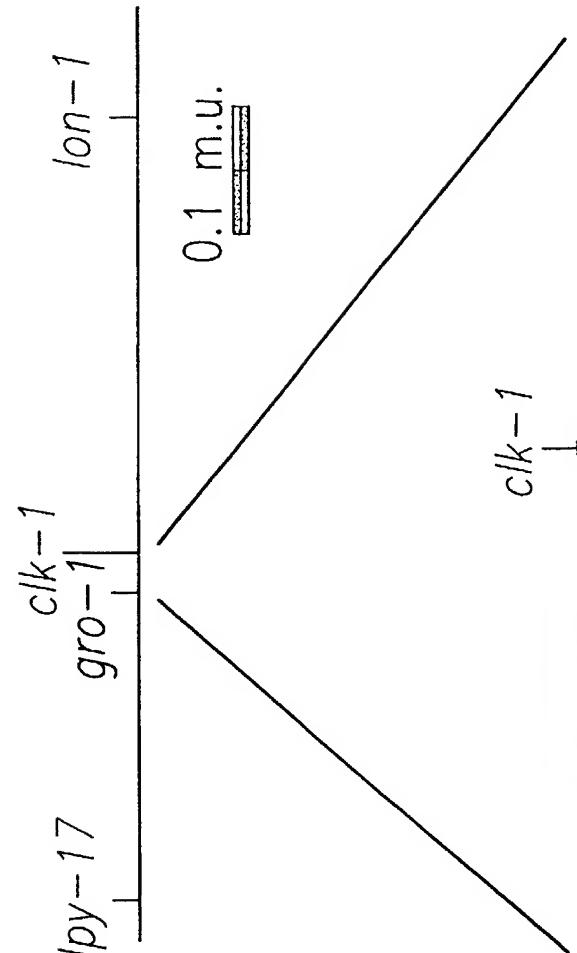

dpy-17
gro-1
clk-1
lon-1
0.1 m.u.

Fig-1A

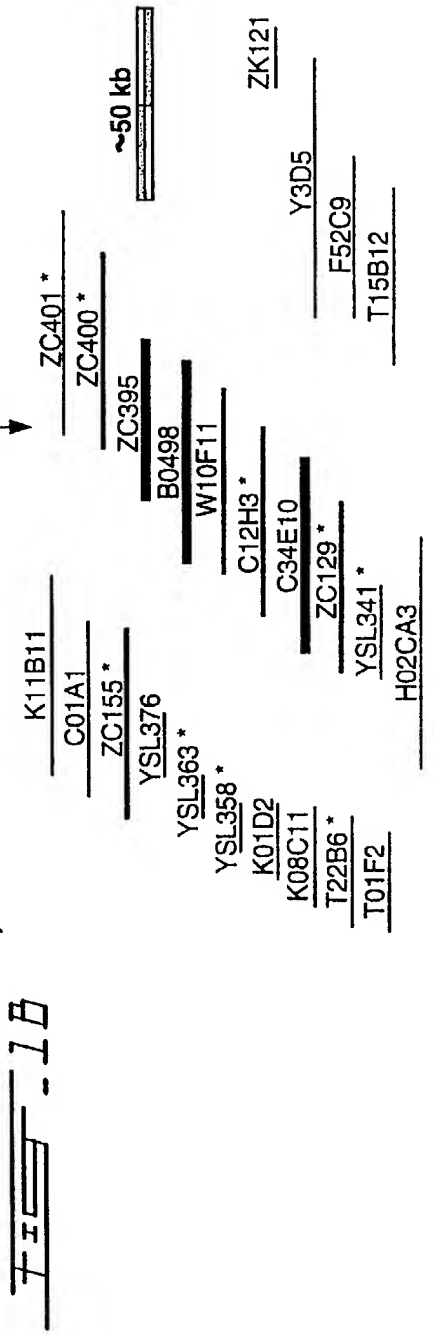

clk-1
K11B11
C01A1
ZC155*
YSL376
YSL363*
YSL358*
K01D2
K08C11
T22B6*
T01F2
ZC401*
ZC400*
ZC395
B0498
W10F11
C12H3*
C34E10
ZC129*
YSL341*
H02CA3
~50 kb
ZK121
Y3D5
F52C9
T15B12

Fig-1B

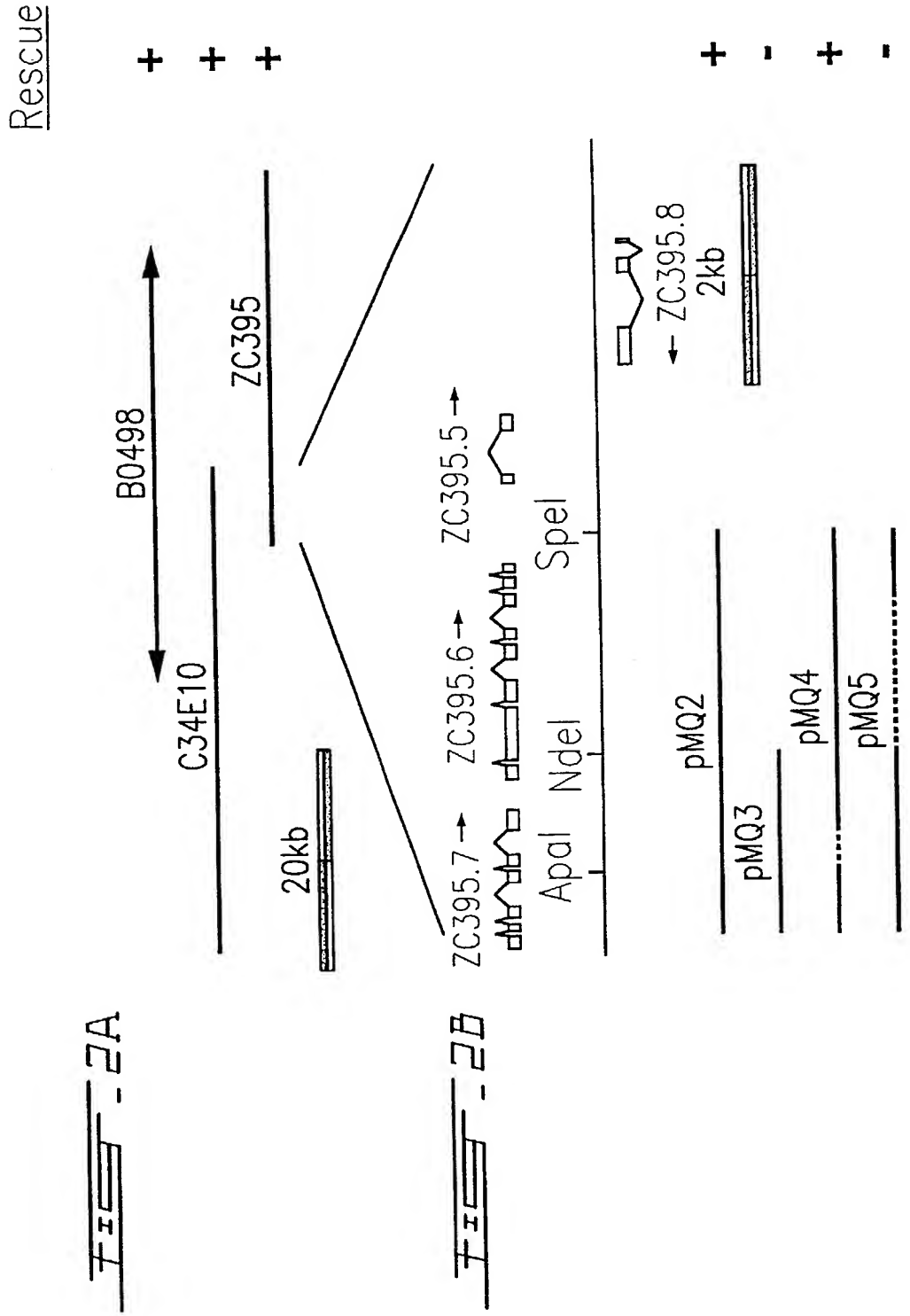

Rescue
B0498
C34E10
ZC395
20kb
ZC395.7
ZC395.6
ZC395.5
ApaI
NdeI
SpeI
ZC395.8
2kb
pMQ2
pMQ3
pMQ4
pMQ5

Fig - 2A

Fig - 2B

gro-1

SL2

M I F R K F L N F L K P Y K M R 16

aaaatatcgtcaggaaataataacatttcagatataccctgaactctacagtttATGATATTCAGGAAATTTCTGAATTTTCTGAAACCTTACAAAATGC 1394

T D P I I F V I G C T G T G K S D L G V A I A K K Y G G E V I S V 49

GAACGGATCCGATTATTTTCGTGATTGGGTGCACTGGAACCGGGAAAAGTGATCTTGGAGTGGCAATTGCAAAGAAATATGGAGGAGAGGTGATTAGTGT 1494

SHP109

D S M Q F Y K G L D I A T N K I T 66

AGATTCAATGCAATTTTATAAAGgtacatgggttttgtttcaattttaaattaattaattttcgtttttcagGACTTGACATTGCCACGAATAAGATAAC 1594

E E E S E G I Q H H M M S F L N P S E S S S Y N V H S F R E V T L 99

GGAAGAAGAATCTGAAGGGATTCAACATCATATGATGTCATTTTTGAATCCATCTGAATCATCATCTTATAATGTACATAGTTTCCGAGAAGTCACGTTG 1694

SHP94

D L I K K I R A R S K I P V I V G 116

GATCTTATTAAAgtgcttaattcgccacttttttgaacttgatcctaattttcataatttttcagAAAATCCGCGCCCGTTCAAAAATTCCTGTAATTGTCG 1794

SHP95

G T T Y Y A E S V L Y E N N L I E T N T S D D V D S K S R T S S E 149

GAGGAACCACTTATTATGCTGAAAGTGTCCTTTATGAGAATAATCTGATTGAAACCAACACTTCAGATGACGTGGATTCCAAATCGAGAACATCATCAGA 1894

SHP96

S S S E D T E E G I S N Q E L W D E L K K I D E K S A L L L H P N 182

ATCGTCATCTGAAGACACTGAAGAAGGAATTAGTAATCAAGAATTATGGGATGAATTGAAAAAAATCGACGAAAAATCAGCACTTCTTCTACATCCAAAT 1994

Fig 3A

gro-1 continued... 4/32

```
N R Y R V Q R A L Q I F R E T G                                          198
AATCGTTATCGAGTACAGAGAGCATTGCAAATTTTCAGAGAAACTGgtaattgatttgcaaatttccagattaaaaacaaatcaagtaaagttttttgca  2094

 I R K S E L V E K Q K S D E T V D L G G R L R F D N S L V I F M D  231
gGAATCCGAAAAAGTGAACTTGTTGAAAAACAGAAATCAGATGAAACTGTTGATTTGGGTGGACGACTACGATTTGATAATTCTTTAGTTATTTTTATGG  2194
SHP97

 A T P E V L E E R L D G R V D K M I K L G L K N E L I E F Y N E   263
ATGCAACACCTGAAGTTTTAGAAGAAAGACTTGATGGAAGAGTTGATAAAATGATTAAATTGGGTTTGAAGAATGAATTGATCGAGTTTTATAACGAGgt  2294

aaatatttgaatttttccagaaaaaaaaagaaaatttttattattttgttttttttttcattctttactattttccaaaaaagtttaaacttttgaaaac  2394

                                                        H A E Y   267
tgttcagaaaatgttcgtgtatttatttagcttactgaggcattatttcattgtgatttttactatactctataaactaaattttcagCACGCCGAGTA  2494

 I N H S K Y G V M Q C I G L K E F V P W L N L D P S E R D T L N G  300
CATAAATCACAGCAAATATGGTGTCATGCAATGTATTGGTCTTAAAGAATTCGTTCCATGGCTCAATTTGGACCCATCAGAAAGAGATACACTCAATGGG  2594
                          CG  e2400 lesion                                  SHP98

D K L F K Q G                                           C D D V K L H T R Q Y  318
GATAAATTGTTCAAGCAAGGgtaatttaaatttattttcaatttttataaattccaagctattttcagATGCGATGATGTGAAGCTTCACACTCGACAAT  2694
```

Fig 3B

gro-1 continued... 5/32

```
  A  R  R  Q  R  R  W  Y  R  S  R  L  L  K  R  S  D  G  D  R                    33

ATGCACGGCGCCAGAGACGGTGGTATCGATCGAGACTTTTAAAACGGTCGGATGGTGATCGGgtatgttgattttaaaaaaattgaatttttaaagaact  279
          SHP99

tttttactaaattaacaaagttattggctgaaaatggctgaaaattatagtaaaactaatcaaaaaaattgaaattttgaattaaagtcataaagtgacg  289

                                                                  K  M  A  S  T  K  M  L  D   34

accagaaaattaaaaaaaaacatttttctattttaattaattcactctacttcactttaaaaataattttcagAAAATGGCAAGTACAAAAATGCTGGAT  299

  T  S  D  K  Y  R  I  I  S  D  G  M  D  I  V  D  Q  W  M  N  G  I  D  L  F  E  D   37

ACATCTGACAAGTACCGAATAATTAGTGATGGAATGGACATTGTTGATCAATGGATGAATGGAATCGATCTATTTGAAGATgtaaaatttcacaaattct  309

                                    I  S  T  D  T  N  P  I  L  K  G  S  D  A  N  I  L  L  N  C  E  I   39

aaaatttccgaatcacaaattaaaatttctacagATCTCCACAGACACCAATCCAATTCTAAAAGGGTCCGATGCAAATATTCTGCTGAATTGTGAAATC  319

  C  N  I  S  M  T  G  K  D  N  W                                    Q  K  E  I  D  G  K  K   41

TGTAATATTTCAATGACTGGAAAAGATAATTGgtttgtttcaatacatattataatttcgaaatgaatttttttcagGCAGAAACATATCGATGGGAAAAA  329
     SHP110      SHP100

  H  K  H  H  A  K  Q  K  K  L  A  E  T  R  T  •                              43

GCACAAGCATCATGCTAAGCAAAAGAAATTGGCAGAGACTCGCACAtaagacgctatatttatttttttgttaacttaaattatttttgttgttgattgtt  339

                                    polyA
                                   /
ctctaaataaaaaaacagctcagagagaagattaggcgctcgtccacatctccgacgatagtcaacccgaacgaagggaactatctttaattgtcagtga  349
                                                                  SHP92
```

Fig 3C

```
tgatttttactatactctataaactaaattttcagCACGCCGAGTACATAAATCACAGCAAATATGGTGTCACG      1197
                                  H  A  E  Y  I  N  H  S  K  Y  G  V  T        276

  TTGGTCTTAAAGAATTCGTTCCATGGCTCAATTTGGACCCATCAGAAAGAGATACACTCAATGGGGATAAATTGT  1272
  L  V  L  K  N  S  F  H  G  S  I  W  T  H  Q  K  W  I  H  S  M  G  I  N  C    301

TCAAGCAAGGgtaatttaaatttatttttcaatttttataaattccaagctattttcagATGCGATGATGtgaagcttc 1350
S  S  K  D                                                 A  M  M  *           308
```

Fig 30

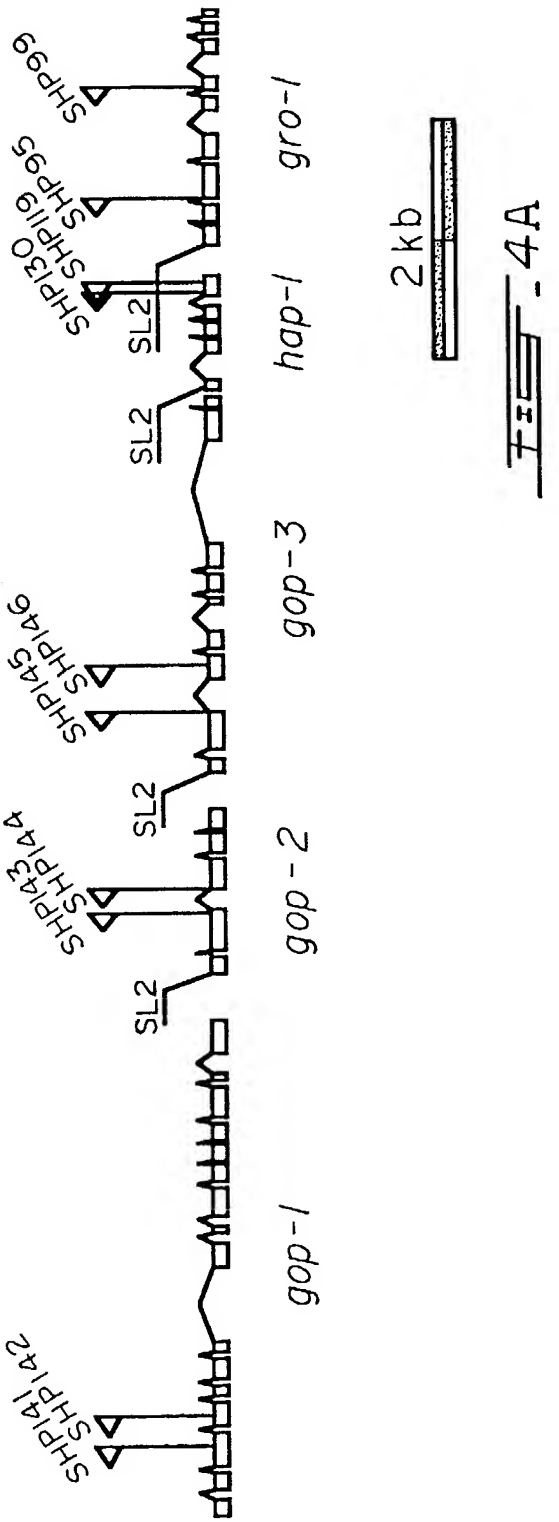

SHpl99
SHpl195
SHpl130
SL2
SL2
gro-1
hap-1
gop-3
gop-2
gop-1
SHpl146
SHpl145
SL2
SHpl144
SHpl143
SL2
SHpl142
2 kb

Fig-4A

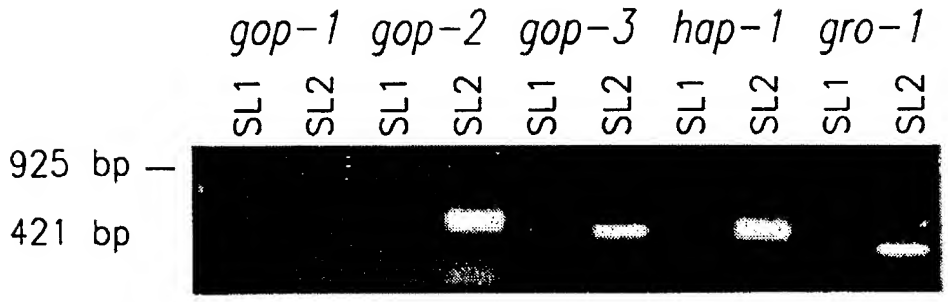

gop-1
gop-2
gop-3
hap-1
gro-1
SL1
SL2
SL1
SL2
SL1
SL2
SL1
SL2
SL1
SL2
925 bp
421 bp

Fig 4B

Sequence of GRO-1 and homologues

```
C.elegans      1  MIFRKFLNFLKPYKMRTDPIIFVIGCTGTGKSDLGVAIAKKYGGEVISVDSMQFYKGLDIATNKITEEESEGIQ
S.cerevisiae   1      MLKGPLKGCLNMSKKVIVIAGTTGVGKSQLSIQLAQKFNGEVINSDSMQVYKDIPIITNKHPLQEREGIP
E.coli         1          MSDISKASLPKAIFLMGPTASGKTALAIELRKILPVELISVDSALIYKGMDIGTAKPNAEELLAAP
```

ATP/GTP binding site

```
C.elegans     76  HMMSFLNPSESSSYNVHSFREVTLDLIKKIRARSKIPVIVGGTTYYAESVLYENNLIETNTSDDVDSKSRTSSE
S.cerevisiae  72  HVMNHVDWSE--EYYSHRFETECMNAIEDIHRRGKIPIVVGGTHYYLQTLFNKRVDTKSSERKLTRKQLDILES
E.coli        68  RLLDIRDPSQ--AYSAADFRRDALAEMADITAAGRIPLLVGGTMLYFKALLEGLSPLPSADPEVRARIEQQAAE
```

```
C.elegans    151  SSEDTEEGISNQELWDELKKIDEKSALLLHPNNRYRVQRALQIFRETGIRKSELVEKQKSDETVDLGGRLRFDN
S.cerevisiae 147  DPDV---------IYNTLVKCDPDIATKYHPNDYRRVQRMLEIYYKTGKKPSETFNEQK--------ITLKFD-'
E.coli       143  GWES---------LHRQLQEVDPVAAARIHPNDPQRLSRALEVFFISGKTLTELTQTSG--------DALPYQV
```

FIG. 5A

e2400

```
C.elegans       226 LVIFMDATPEVLEERLDGRVDKMIKLGLKNELIEFYNEHAEYINHSKYGVMQCIGLKEFVPWLNLDPSERDTLN
S.cerevisiae    205 LFLWLYSKPEPLFQRLDDRVDDMLERGALQEIKQLYEYYSQNKFTPEQCENGVWQVIGFKEFLPWLTGKTDDNT
E.coli          202 QFAIAPASRELLHQRIEQRFHQMLASGFEAEVRALFARGDLHTDLPSIRCVGYRQMWSYLEGEISYDEMVYRGV

C.elegans       301 DKLFKQGCDDVKLHTRQYARRQRRWYRSRLLKRSDGDRKMASTKMLDTSDKYRIISDGMDIVDQWMNGIDLFED
S.cerevisiae    280 KLEDCIERMKT--RTRQYAKRQVKWIKKMLIPDIKGDILLDATDLSQWDTNASQRAIAISNDFISNRPIKQERA
E.coli          277 ------------ATRQLAKRQITWLRGWEGVHWLDSEKPEQARDEVLQVVGAIAG

                                C2H2 zinc finger
C.elegans       376 STDTNPILKGSDANILLNCEICNISMTGKDNWQKHIDGKKHKHHAKQKKLATRT
S.cerevisiae    353 KALEELLSKGETTMKKLDDWTHYTRNVCRNADGKNVVAIGEKYWKIHLGSRRHKSNLKRNTRQADFEKWKINKK
```

Fig 5B

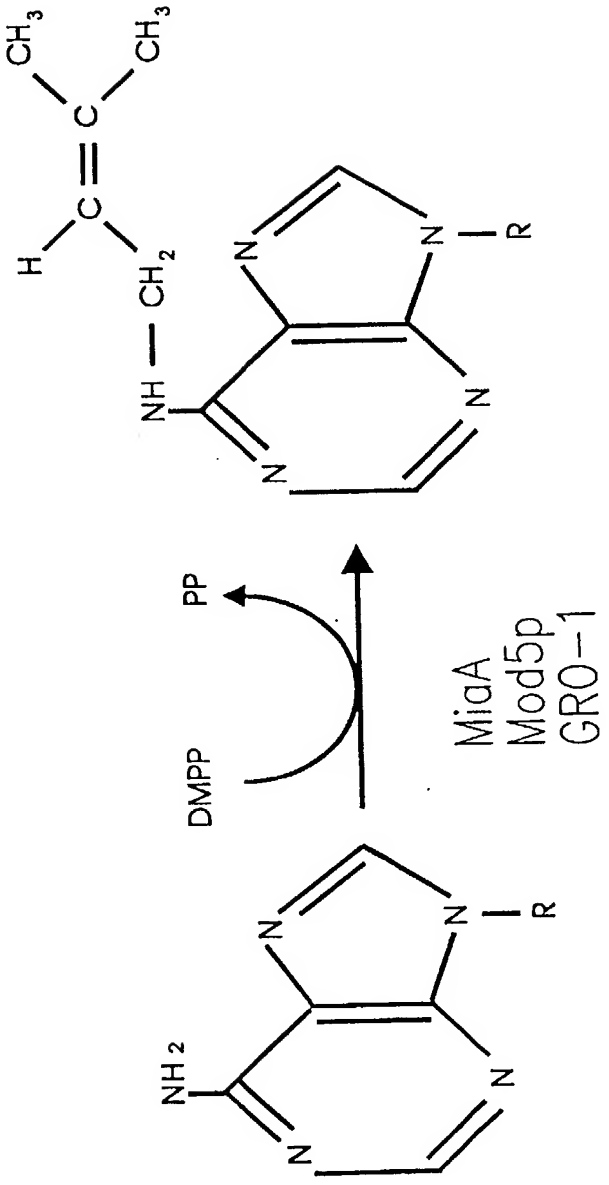


FIG - 6

Sequence of HAP-1 and its homologues

H. sapiens	MAASLVGKKIVFVTGNAKKLEEVVQILGDKFP-----CTLVAQKIDLPEYXG-EPDEISIQKCQE
C. elegans	MLYILWKLNYLQKKMSLRKINFVTGNVKKLEEVKAILKNFE--------VSNVDVDLDEFQG-EPEFIAERKCRE
S. cerevisiae	MSNNEIVFVTGNANKLKEVQSILTQEVDNNNKTIHLINEALDLEELQDTDLNAIALAKGKQ
E. coli	MQKVVLATGNVGKVRELASLLSDFGLD-----IVAQTDLGVDSAEETGLTFIENAILKA

H. sapiens	AVRQV-QG-PVLVEDTCLCFNALGXLPGPYIKWFL--EKLKPEGLHQLLAGFED-----KSAYALCTFALSTGDP
C. elegans	AVEAV-KG-PVLVEDTSLCFNAMGGLPGPYIKWFL--KNLKPEGLHNMLAGFSD-----KTAYAQCIFAYTEG-L
S. cerevisiae	AVAALGKGKPVFVEDTALRFDEFNGLPGAYIKWFL--KSMGLEKIVKMLEPFEN-----KNAEAVTTICFADSRG
E. coli	RHAAKVTALPAIADDSGLAVDVLGGAPGIYSARYSGEDATDQKNLQKLLETMKDVPDDQRQARFHCVLVYLRHAE

H. sapiens	SQPVRLFRGRTSGRIV-APRGCQDFGWDPCFQP-DGYEQTYAEMPKAEKNAVSHRFRALLELQEYFGSLAA
C. elegans	GKPIHVFAGKCPGQIV-APRGDTAFGWDPCFQP-DGFKETFGEMDKDVKNEISHRAKALELLKEYFQNN
S. cerevisiae	E--YHFFQGITRGKIV-PSRGPTTFGWDSIFEPFDSHGLTYAEMSKDAKNAISHRGKAFAQFKEYLYQNDF
E. coli	DPTPLVCHGSWPGVITREPAGTGGFGYDPIFFV-PSEGKTAAELTREEKSAISHRGQALKLLLDALRNG

FIG. 7

mRNA sequence of human homologue of *gro-1*: hgro-1

CTGCCATAAG **ATG**GCGTCCG TGGCGGCTGC ACGAGCAGTT CCTGTGGGCA
GTGGGCTCAG GGGCCTGCAA CGGACCCTAC CTCTTGTAGT GATTCTCGGG
GCCACGGGCA CCGGCAAATC CACGCTGGCG TTGCAGCTAG GCCAGCGGCT
CGGCGGTGAG ATCGTCAGCG CTGACTCCAT GCAGGTCTAT GAAGGCCTAG
ACATCATCAC CAACAAGGTT TCTGCCCAAG AGCAGAGAAT CTGCCGGCAC
CACATGATCA GCTTTGTGGA TCCTCTTGTG ACCAATTACA CAGTGGTGGA
CTTCAGAAAT AGAGCAACTG CTCTGATTGA AGATATATTT GCCCGAGACA
AAATTCCTAT TGTTGTGGGA GGAACCAATT ATTACATTGA ATCTCTGCTC
TGGAAAGTTC TTGTCAATAC CAAGCCCCAG GAGATGGGCA CTGAGAAAGT
GATTGACCGA AAAGTGGAGC TTGAAAAGGA GGATGGTCTT GTACTTCACA
AACGCCTAAG CCAGGTGGAC CCAGAAATGG CTGCCAAGCT GCATCCACAT
GACAAACGCA AAGTGGCCAG GAGCTTGCAA GTTTTTGAAG AAACAGGAAT
CTCTCATAGT GAATTTCTCC ATCGTCAACA TACGGAAGAA GGTGGTGGTC
CCCTTGGAGG TCCTCTGAAG TTCTCTAACC CTTGCATCCT TTGGCTTCAT
GCTGACCAGG CAGTTCTAGA TGAGCGCTTG GATAAGAGGG TGGATGACAT
GCTTGCTGCT GGGCTCTTGG AGGAACTAAG AGATTTTCAC AGACGCTATA
ATCAGAAGAA TGTTTCGGAA AATAGCCAGG ACTATCAACA TGGTATCTTC
CAATCAATTG GCTTCAAGGA ATTTCACGAG TACCTGATCA CTGAGGGAAA
ATGCACACTG GAGACTAGTA ACCAGCTTCT AAAGAAAGGA CCTGGTCCCA
TTGTCCCCCC TGTCTATGGC TTAGAGGTAT CTGATGTCTC GAAGTGGGAG
GAGTCTGTTC TTGAACCTGC TCTTGAAATC GTGCAAAGTT TCATCCAGGG
CCACAAGCCT ACAGCCACTC CAATAAAGAT GCCATACAAT GAAGCTGAGA
ACAAGAGAAG TTATCACCTG TGTGACCTCT GTGATCGAAT CATCATTGGG
GATCGCGAAT GGGCAGCGCA CATAAAATCC AAATCCCACT TGAACCAACT
GAAGAAAAGA AGAAGATTGG ACTCAGATGC TGTCAACACC ATAGAAAGTC
AGAGTGTTTC CCCAGACTAT AACAAAGAAC CTAAAGGGAA GGGATCCCCA
GGGCAGAATG ATCAAGAGCT GAAATGCAGC GTTTAAGAGA CATGTCCAGT
GGCCTTTGGA AAGGTGGTGG GGATCCAGTT CAGGAGGGAG GGGTATGTTT
GTCTCCCAGT CTGGGCAAAG GAGTGCTATG CGGAATTCTC TGCATAGCAG
AAAAGCTCCC ACCATTTTCT TTTGATGTGG TTTTAAAGTC TCACGTTCTC
TATAATAGAA ACAGCAGGTC TTGTCAGCTC CTTGTGTGGC TGATGTGTCT
GGAAATGATG TAGTTCAGGA AAGCATTTTT TTTTTCTTTG AACCTTAAAG
GTTCTATTAT TAAAAGCAGC ACAGATTCCA CATTTTTATA CATGAGGATC
TTCTTTGTGG TGAATACCAG GATTGACTGC ATCCCTTTAA AAGAAGTTTT
ATGTCCCTGA CTCTGGCTAA AATTATCTAA TTTCCAGATG CTTTTGTAGA
TGACTGAAGT ATTTGTGAGC CACATATTGG GAGTTCTAGA TTTGAGTGAA
TGGCAGGAAA GGGCCATCTC CATTGAGATG ATTAAGTGAA CCAAACTAGT
TCTCGGAATT CTACAGAGAA GGAGGGAATC AGACTGAGGA AGCTGTGACA
TAGGACTTGA AGACCAAAGA CTTTGAAATT TGCGAGCTGC TCATGTGTGA
GTTATTATCA CTGCTGTCTT TCTATTGAGT TACAAATCTA TATTTTTATT
GAAGTTTAAA TAAAGAAAAA ATTTACAAGA AAAAAAAAAA A

FIG. 8

GRO-1 and its human homologue hgro-1p

```
                                  ..  .    . ......  .       ...  . ....  . ....  ..
hgro-1p          MASVAAARAVPVGSGLRGLQRTLPLVVILGATGTGKSTLALQLGQRLGGEIVSADSMQVYEGLDIITN
GRO-1              MIFRKFLNFLKPYKMRTDPIIFVIGCTGTGKSDLGVAIAKKYGGEVISVDSMQFYKGLDIATN

                   .   .    ... ..  .    . .  ..    ..  . .. ...  .... .. .. .
hgro-1p          KVSAQEQRICRHHMISFVDPL-VTNYTVVDFRNRATALIEDIFARDKIPIVVGGTNYYIESLLWKVLVN
GRO-1            KITEEESEGIQHHMMSFLNPSESSSSYNVHSFREVTLDLIKKIRARSKIPVIVGGTTYYAESVLYENNLI

                              .   ..       .   .   .   .  ...    . . .. . ....
hgro-1p          TKPQEMGTEKVIDRKVELEKEDGLV------LHKRLSQVDPEMAAKLHPHDKRKVARSLQVFEETGISH
GRO-1            ETNTSDDVDSKSRTSSESSSEDTEEGISNQELWDELKKIDEKSALLLHPNNRYRVQRALQIFRETGIRK
```

FIG 9A

```
                 ..    .  .   ... . .        .  .. .... ... .  ..  ..     .
hgro-1p          SEFLHRQHTEEGGGPLGGPLKFSNPCILWLHADQAVLDERLDKRVDDMLAAGLLEELRDFHRRYNQKNV
GRO-1            SELVEKQKSDETVD-LGGRLRFDNSLVIFMDATPEVLEERLDGRVDKMIKLGLKNELIEF---YNEHAE

                       .  .  . .. ...   .           . .
hgro-1p          SENSQDYQHGIFQSIGFKEFHEYLITEGKCTLETSNQLLKKGPGPIVPPVYGLE----------
GRO-1            YINHSKY--GVMQCIGLKEFVPWLNLDPSERDTLNGDKLFKQGCDDVKLHTRQYARRQRRWYRSRLLK

                    ..     .          .              .
hgro-1p          VSDVSKWEESVLEPALEIVQSFIQGHKPTATPIKMPYNEAENKRSYHL----------
GRO-1            RSDGDRKMASTKMLDTSDKYRIISDGMDIVDQWMNGIDLFEDISTDTNPILKGSDANILLN

                    .  .    .  .  ..  . .   .
hgro-1p          CDLCDRIIIGDREWAAHIKSKSHLNQLKKRRRLDSDAVNTIESQSVSPDYNKEPKGKGSPGQNDQELKCSV
GRO-1            CEICNISMTGKDNWQKHIDGKKHKHHAKQKKLAETRT

                 C2H2 zinc finger
```

FIG 9B

Conceptual translation of a partial sequence of the Drosophila homologue of *gro-1*

P I T C K H K K Q L T A T S G S V P I G I H V L K T C G F Y L P *Stop* L T *Stop* I H S Q *Stop* V E
Met I R K V P L I V V L G S T G T G K T K L S L Q L A E R F G G E I I S A D S **Met** Q V V Y T H L
D I A T A K A T K E E Q S R A R H H L L D V A T P A E P F T V T H F R N A A L P I V E R L L
A K D T S P I V V G G T N Y Y I E S L L W D I L V D S D V K P D E G K H S G E H L K D A E L
N A L S T L E L H Q H L A K I D A G S A N R I H P N N R R K I I R A I E V Y Q S T G Q T

Fig - 10

17/32

Structure of pMQ8

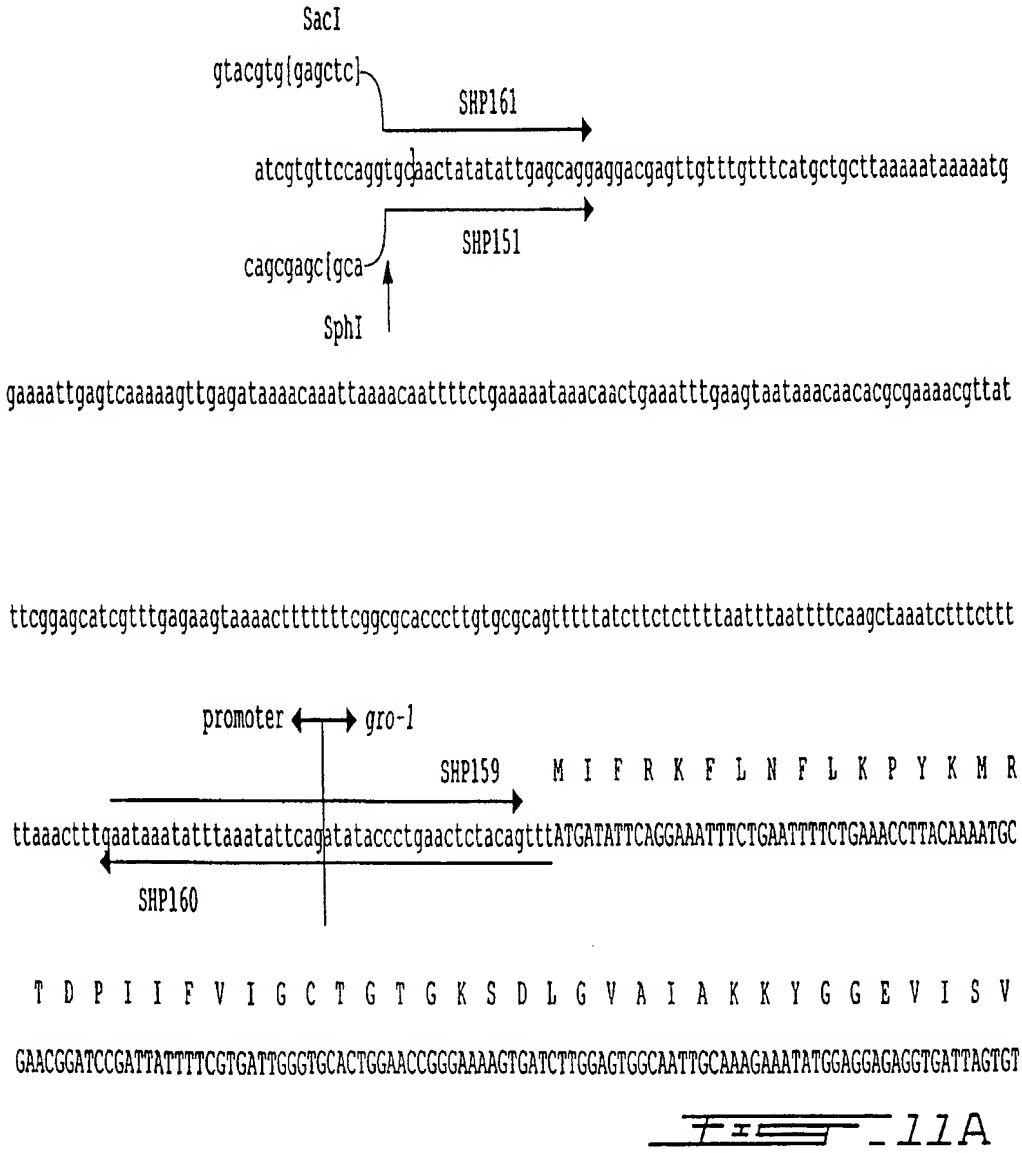


Fig 11A

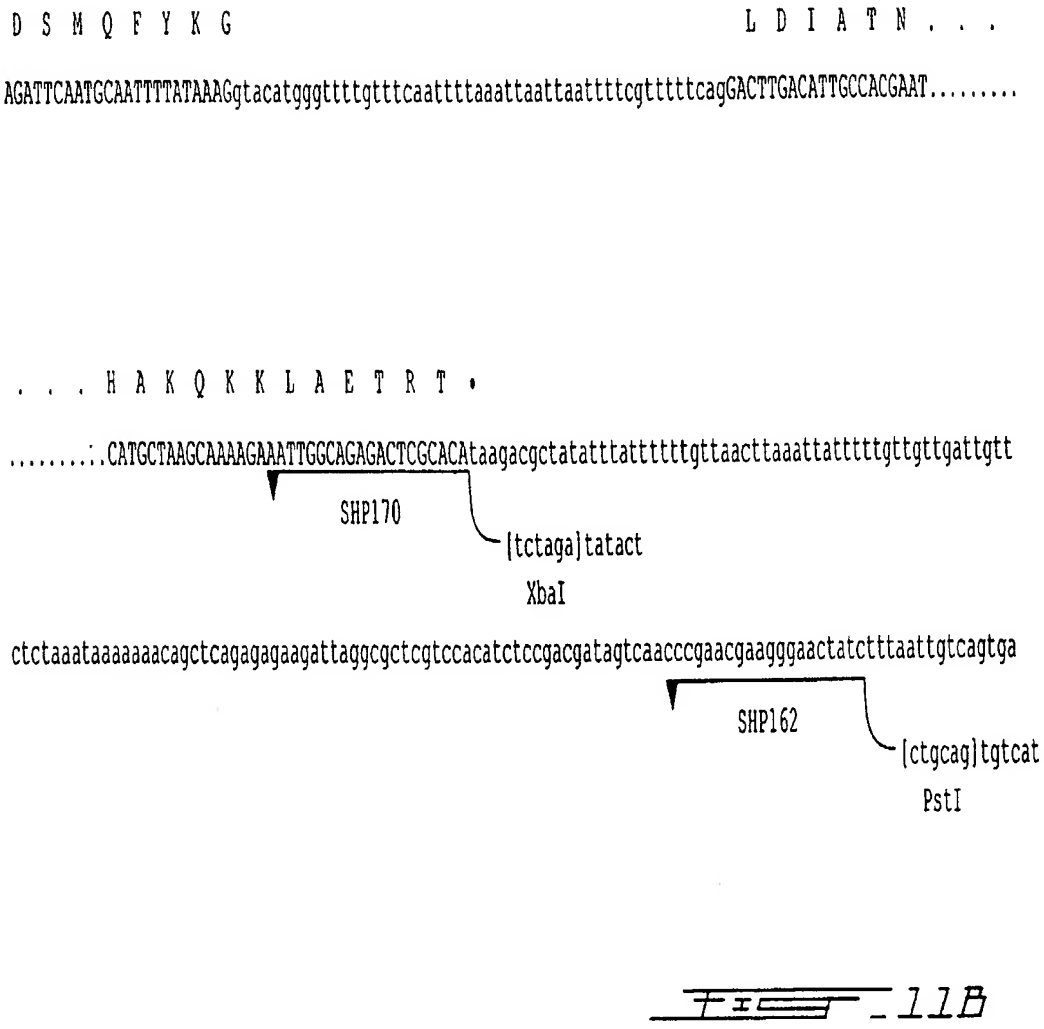

D S M Q F Y K G
L D I A T N . . .
AGATTCAATGCAATTTTATAAAGgtacatgggttttgtttcaattttaaattaattaattttcgtttttcagGACTTGACATTGCCACGAAT.....
. . . H A K Q K K L A E T R T •
.....CATGCTAAGCAAAAGAAATTGGCAGAGACTCGCACAtaagacgctatatttatttttgttaacttaaattatttttgttgttgattgtt
SHP170
[tctaga]tatact
XbaI
ctctaaataaaaaaacagctcagagagaagattaggcgctcgtccacatctccgacgatagtcaacccgaacgaagggaactatctttaattgtcagtga
SHP162
[ctgcag]tgtcat
PstI

Fig. 11B

Construction of pMQ18

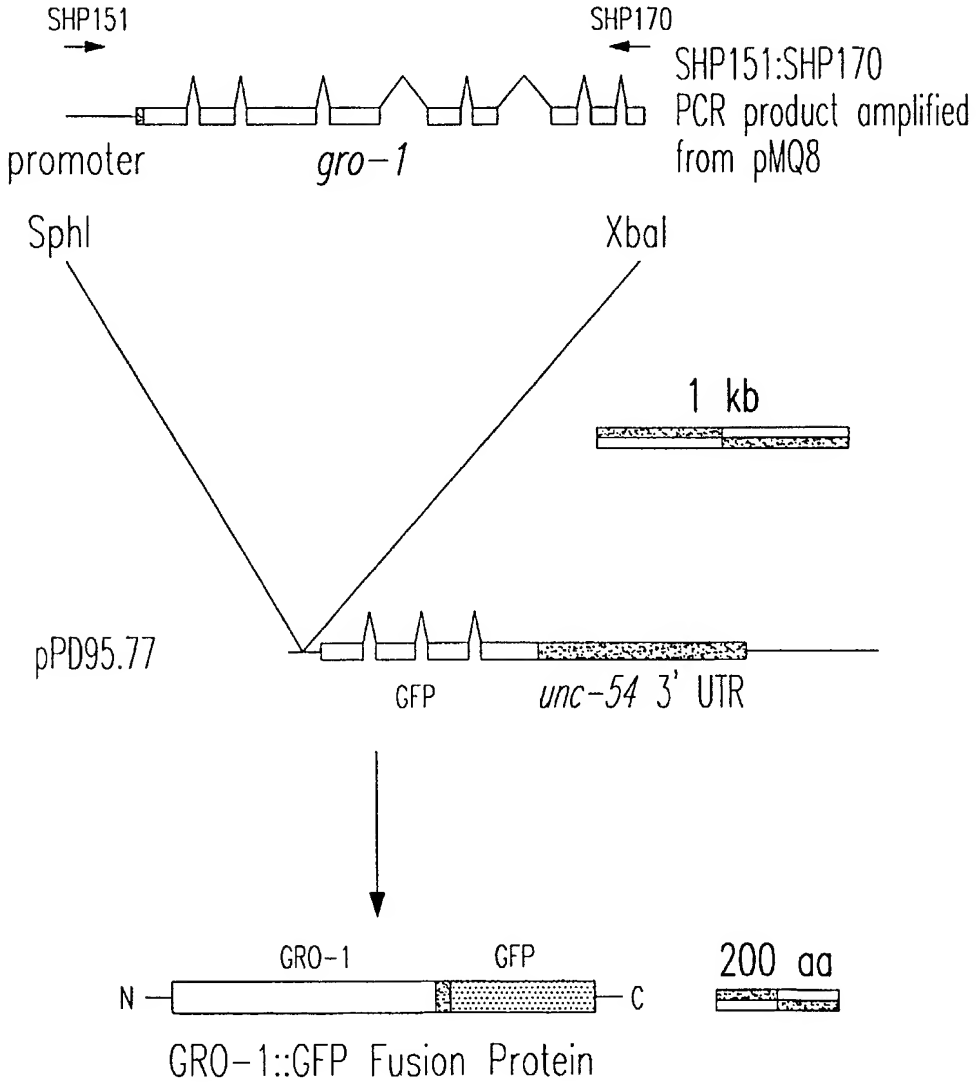


Fig. 12

gop-1

```
atcgtgttccaggtgcaactatatattgagcaggaggacgagttgtttgtttcatgctgcttaaaaataaaaatggaaaattgagtcaaaaagttgagat  -9557

aaaacaaattaaaacaattttctgaaaaataaacaactgaaatttgaagtaataaacaacacgcgaaaacgttatttcggagcatcgtttgagaagtaaa  -9457

actttttttcggcgcacccttgtgcgcagtttttatcttctcttttaatttaattttcaagctaaatctttctttttaaactttgaataaatatttaaat  -9357

                                                  M  F  R  K  L  G  S  S  G  S  L  W  K  P  K  N  P  H  S  L  E    21
attcagaatgcaccaataaacctggaacaaaatcgataATGTTCCGCAAGCTTGGTTCTTCTGGGTCACTATGGAAGCCGAAAAATCCGCATTCTTTGGA  -9257
                                                                  SHP190

Y  L  K  Y  L  Q  G  V  L  T  K  N  E  K  V  T  E  N  N  K  K  I  L  V  E  A  L  R  A  I  A  E  I    54
ATACCTCAAATATTTACAAGGAGTGCTCACAAAAAATGAGAAAGTTACGGAAAACAATAAGAAAATATTAGTAGAAGCATTACGAGCTATCGCAGAAATT  -9157

L  I  W  G  D  Q  N  D  A  S  V  F  D                                                F  F  L  E  R    72
CTCATTTGGGGCGATCAGAATGATGCTTCGGTTTTTGAgtgagttttttccaatgttttttttcaaatctgatgttgaatttcagTTTCTTCCTTGAGC  -9057

  Q  N  L  L  Y  F  L  K  I  M  E  Q  G  N  T  P  L  N  V  Q  L  L  Q  T  L  N  I  L  F  E  N  I  R   105
GGCAAATGCTTCTTTATTTCTTGAAAATTATGGAACAAGGAAACACACCACTAAATGTACAATTACTGCAGACTTTGAACATTTTATTCGAAAATATTCG  -8957
                                                                  SHP171

H  E  T  S  L  Y                                                F  L  L  S  N  N  H  V  N  S  I  I   123
ACATGAAACTTCACTTTgtaagttttttatatggattttcgcttaaaattgccagttttcagATTTCCTTCTAAGTAACAATCATGTAAACTCGATTATT  -8857

S  H  K  F  D  L  Q  N  D  E  I  M  A  Y  Y  I  S  F  L  K  T  L  S  F  K  L  N  P  A  T  I  H  F  F   157
TCCCACAAATTCGATTTACAAAATGATGAGATCATGGCTTACTACATTAGTTTTCTGAAAACTCTTTCATTTAAACTGAATCCAGCTACAATCCACTTCT  -8757
```

FIG 13A

gop-1 continued...

```
F N E T T E E F P L L V E V L K L Y N W N E S M V R I A V R N I L   190
TCTTCAATGAAACGACTGAAGAATTTCCATTGTTGGTAGAAGTTTTGAAGCTTTATAATTGGAATGAATCAATGGTTCGAATTGCTGTTAGAAATATTCT  -8657
                  SHP141                    SHP172
L N I V R V Q D D S M I I F A I K H T K                           210
TTTAAATATTGTGAGAGTTCAAGATGATTCAATGATTATTTTCGCTATCAAGCATACAAAAgttagtagaaaattattttgaaaaggtgtatttaagcaa  -8557

        E Y L S E L I D S L V G L S L E M D T F V R S A E N V L A N   240
taaatattacagGAATATCTATCGGAGTTAATAGATTCTCTAGTTGGTCTCTCACTTGAAATGGACACATTTGTACGATCTGCTGAGAATGTGTTAGCTA  -8457

R E R L R G K V D D L I D L I H Y I G E L L D V E A V A E S L S I   273
ATCGAGAGAGATTACGAGGAAAAGTGGATGATTTAATTGATTTGATTCATTATATTGGTGAACTATTGGATGTGGAAGCTGTCGCCGAAAGTTTATCAAT  -8357
SHP142                    SHP173
L V                                 T T R Y L S P L L L S S I S P R   291
TTTAGgtcagttttactgctggaaaatcaagtttttaatgttaaattttcagTAACAACACGATACTTAAGCCCTCTATTACTTTCAAGTATATCACCAA  -8257

R D N H S L L L T P I S A L F F F S E F L L                       313
GAAGAGATAATCATTCACTTCTACTCACTCCGATTTCTGCGTTATTTTTTTTCTCTGAATTTTTATTGgtgagttttaacatttaaaattacatttttct  -8157

            I V R H H E T I Y T F L S S F L F D T Q N T L T T H W I   341
aatttatttatttttcagATAGTTCGTCACCATGAAACAATATATACATTTTATCATCTTTCCTATTTGACACTCAGAATACTTTGACGACCCATTGGA  -8057

R H N E K Y C L E P I T L S S P T G E Y V N E D H                  366
TACGTCATAATGAGAAATATTGCTTAGAACCGATTACATTATCATCACCAACCGGAGAATATGTGAATGAAGACCAgtaagagctgaaattttaaaattt  -7957

                V F F D F L L E A F D S S Q A D D S K A F Y G L M   391
ttgctttgaatatagtattttcagCGTATTTTTCGATTTTCTACTGGAAGCATTTGATTCCAGTCAAGCAGACGATTCGAAGGCATTCTATGGATTAATG  -7857
```

Fig 13B

gop-1 continued...

```
L  I  Y  S  M  F  Q  N  N  A                                                    401
CTGATTTATTCAATGTTTCAGAATAATGgtgagttttaaaaaattgatttgttaaattaaaatttccatttccaataactcctcttcagacagtaagttt -7757

tcaatgttgtaaagttcctgttcatctgtgatcgttttcttcatttttttagttttgcatgaacagttttcaaatttttttgatatcatacagtaaatat -7657

cgtcatccagataattttctatttaaaaaaaatgaataaaaagagggcgcgcagaaattgccgaagtaatgtaaatttaaagggacacatgcgtagcttg -7557

ttgtgtgggtctcgccgcgctttgtttgatttatcttgttttctgctcaaagagctgtttttatttagcgttgaatgctttttaccgttctcatcggc -7457

tttttaataggaatatttaaaaaaaaaggtttaataaatcttcgtttttacaaaatccatctaagatttgcatttgtgaagctcaacaagtaaagtttta -7357

agtaacattgtttttttaaaaaacaattgaaccaaattttgccgaaacattaataacatgacgatactctataaaatattcctctttttcaaaataaatttt -7257

                                  D  V  G  E  L  L  S  A  A  N  F  P  V  L  K  E  S  T  T  T  S  L  A  Q  Q  N   427
caaaaaaaatccatttttcagCCGATGTTGGAGAACTTCTATCTGCTGCCAACTTCCCAGTGCTCAAAGAATCAACGACAACTTCATTAGCTCAACAGAA -7157
                                                    SHP174

L  A  R  L  R  I  A  S  T  S  S  I  S  K  R  T  R  A  I  T  E  I  G  V  E  A  T  E  E  D  E  I  F   480
TCTTGCTCGTCTCCGAATAGCATCTACGTCTTCCATATCAAAGCGAACGAGAGCTATCACTGAAATTGGAGTAGAAGCGACCGAGGAAGATGAGATTTTT -7057
                                                                SHP185

H  D  V  P  E  E  Q  T  L                                                       469
CATGATGTTCCTGAAGAACAAACGTTGgtaagtaaataaatcaacattgattgttacacaaactttaatatttttaaatttgaaaattttcttcaaagtg -6957

                                  E  D  L  V  D  D  V  L  V  D  T  E  N  S  A  I  S  D  P  E                  489
ctcaaaaatcctgtcgaaaattacagGAAGATCTGGTGGATGATGTATTGGTTGATACTGAAAATTCAGCAATAAGTGATCCAGAAgtgagtagaaaacg -6857

                                                                P  K  N  V  E  S  E  S  R   498
tgcatgtattaattattaaaaaaaaaatatagttttccccagttttccttgacctaaaactcagcaatttcagCCTAAAAACGTGGAGTCAGAATCTCGT -6757
```

FIG - 13C

gop-1 continued... 23/32

```
S R F Q S A V D E L P P P S T S G C D G R L F D A L S S I I K A V G  532
TCTCGATTTCAATCTGCTGTTGATGAGCTTCCACCTCCGTCGACTTCTGGATGTGATGGTCGACTTTTTGATGCACTTTCATCGATTATCAAAGCAGTTG -6657

 T D D N R I R P I T L E L A C L V I R Q I L M T V D D E K          561
GAACAGATGACAATCGAATTCGACCAATTACATTGGAACTTGCATGTCTTGTAATTCGGCAAATTTTAATGACTGTTGATGATGAAAAAgtaagattaca -6557
          SHP175
                                  V H T S L T K L C F E V R L K L L S  579
aattcaaaattgagcaaaatcagaatctaaatttcataaattgttcagGTACATACCAGTTTAACGAAATTATGCTTCGAAGTTCGTCTAAAACTTTTAT -6457

 S I G Q Y V N G E N L F L E W F E D E Y A E F E                    603
CATCAATTGGACAATATGTTAATGGAGAGAATCTGTTTTTGGAGTGGTTTGAGGATGAATATGCAGAATTTGAAgtaagccaagaggtccgaaaataatt -6357

                    V N H V N F D I I G H E M L L P P A A T P L S N L L L  630
taattcatccttttttattcagGTGAATCACGTGAATTTCGATATAATCGGTCACGAAATGCTTCTTCCTCCAGCTGCAACTCCTCTTTCGAATCTGCTAC -6257

 H K R L P S G F E E R I R T                              Q I V  647
TTCATAAGCGATTGCCCAGTGGATTTGAAGAACGAATAAGAACTgtaggaaacttttaaatttgaaaattaattatatatatatttgcagCAAATCGTA -6157

F Y L H I R K L E R D L T G E G D T E L P V R V L N S D Q E P V A I  681
TTCTACCTACATATTCGAAAATTGGAACGAGATTTGACCGGTGAAGGAGACACAGAATTACCTGTGAGAGTGTTGAATTCTGATCAGGAACCAGTTGCCA -6057

 G D C I N L H                                          N S D L L S C T  696
TCGGTGATTGTATTAATTTACgtgagttcatctgcatagaaaacaccatatttctactcaaattaacaattttcagATAATTCGGATCTTCTATCCTGCA -5957

 V V P Q Q L C S L G K P G D R L A R F L V T D R L Q L I L V E P D  729
CTGTGGTTCCTCAACAACTATGTTCTCTTGGAAAACCTGGTGATCGTCTTGCTCGATTCCTTGTCACTGATAGACTTCAATTAATTCTTGTCGAACCGGA -5857
 SHP176
 S R K A G W A I V R F V G L L Q D T T I N G D S T D S K V L H V V  762
TTCTCGAAAAGCCGGATGGGCAATTGTTCGATTCGTAGGACTTCTTCAAGATACAACAATTAATGGAGATTCTACGGATTCGAAAGTTTTGCATGTTGTG -5757
                                                    SHP177
 V E G Q P S R I K                                      K R H P V L T A  779
GTGGAAGGGCAACCCTCGAGAATTAAGgtaagaatactaacgggaaaaaaaaatcaaaaaattacttctgtttcagAAAAGACATCCGGTTTTAACTGCA -5657
```

FIG 13D

gop-1 continued...

```
A F I F D D H I R C M A A K Q R L T K                                    798
AAGTTCATATTCGATGATCACATTCGGTGTATGGCAGCAAAGCAACGGCTCACCAAGgtaacggaaaaaataaccaaaaagacggaaagttattgtaaat -5557

ggacgaaatcggcgaaattaattgaaaacgtttgaatttgccgctaaaaccaaacgaaaaccaaacgaaagcgaaatttaactatcccttcaggtagaat -5457

                    G R Q T A R G L K L Q A I C S A L G V P R I D P A T   824
atacattttatttctctttatagGGTCGCCAAACAGCACGTGGTCTGAAACTTCAGGCGATATGTTCAGCTCTTGGAGTTCCACGTATCGATCCAGCGAC -5357

 M T S S P R M N P F R I V K G C A P G S V R K T V S T S S S S S Q    857
AATGACGTCATCACCACGAATGAATCCATTCAGAATTGTGAAAGGATGCGCACCGGGAAGTGTACGAAAAACTGTTTCCACATCATCATCGTCAAGCCAA -5257

G R P G H Y S A N L R S A S R N A G M I P D D P T Q P S S S S E R R  891
GGACGTCCCGGACATTATTCTGCAAATCTTAGATCAGCATCTAGAAATGCAGGAATGATACCAGATGATCCAACTCAACCGAGTAGTTCTTCGGAAAGAA -5157
                                  SHP178

 S *                                                                    892
GATCCtagggatcaatatctcttcagtttcatcattttatgctgtaaattgtatttaagtattcctattctttgtagtactgtatttacacatcgtctag -5057

ttaaaatcacaaatctccgaaaaaacaaaccagtgaacatgtgatatttctcttgcccatagttctctttttttttttgaaacaaaaacaattacttttat -4957
                                                                  polyA
gctcacctattcgagccatatttttttcccaattaccggttgtttattttaatttctttttttttttctgtaaatctactttatttttaaaactgcatttg -4857

agattgtgtatatttttttcaaaatggttcaaatgccgaatctatctactt                    -4807
```

Fig 13E

gop-2

SL2

M A E K A E N L P S S S A E A S E 1

tttaatcattattcaaacagaaaaaccgattatttattcagattctcaaaaATGGCTGAAAAAGCTGAAAATCTTCCATCTTCTTCGGCCGAAGCTTCAG -470

E P S P Q T G P N V N Q K P S I L V L G M A G S G K T T F V Q 4

AAGAGCCATCACCTCAAACTGGACCAAATGTGAATCAAAAACCATCGATTTTGGTTCTTGGAATGGCTGGTTCTGGAAAAACGACATTTGTTCAGgtaac -460

R L T A F L H A R K T P P Y V I N L D P 6

tttcattcaattttgagagttttcaaacattactattttcagCGTCTCACAGCATTCCTACATGCTCGTAAAACACCTCCATATGTGATTAATCTGGATC -450

A V S K V P Y P V N V D I R D T V K Y K E V M K E F G M G P N G A 10

CGGCAGTTAGCAAAGTACCTTATCCAGTGAATGTTGACATTCGAGATACTGTGAAATACAAGGAAGTTATGAAAGAATTCGGAATGGGACCAAATGGAGC -440

SHP179

I M T C L N L M C T R F D K V I E L I N K R S S D F S V C L L D T 13

AATTATGACATGTCTTAACCTGATGTGTACTCGTTTTGATAAAGTAATTGAGTTGATTAATAAGAGATCTTCTGATTTCTCAGTTTGTCTTCTTGATACT -430

SHP180

P G Q I E A F T W S A S G S I I T D S L A S S H P T 16

CCTGGACAAATTGAAGCATTCACTTGGAGTGCTAGTGGATCTATTATCACTGATTCATTGGCAAGTAGCCATCCCACGgtaagggattttgatttatgaa -420

SHP143

atctgcttgaaatgaaaaaagattctaataaatttttgacttttaaacatttttttacagttatatttggtctattttctatcattaaaagcaaaatgaaa -410

V V M Y I V D S A R A T N P T T F M S N 18

agtcgattctactccatatttattaatttcgacttttcagGTGGTAATGTACATTGTGGATTCCGCTCGTGCCACAAATCCAACTACATTCATGTCCAAT -400

SHP144

FIG 14A

gop-2 continued... 26/32

```
M L Y A C S I L Y R T K L P F I V V F N K A D I V K P T F A L K W M  21
ATGCTCTACGCATGTTCCATTCTCTACCGTACCAAACTTCCATTCATTGTCGTTTTCAACAAAGCTGATATTGTCAAACCAACATTTGCACTCAAATGGA -390

 Q D F E R F D E A L E D A R S S Y M N D L S R S L S L V L D E F Y  24
TGCAAGATTTCGAAAGATTTGATGAAGCTTTAGAGGATGCCAGAAGCAGTTATATGAATGATTTGAGTCGTTCATTGAGTCTCGTTCTTGATGAATTCTA -380
                    SHP181

 C G L K T V                                 C V S S A T G E G F E D V  26
TTGCGGACTGAAAACAGgtttttattcgaaataaaaccttttttaaataataaatttcagTTTGCGTCAGTTCTGCAACTGGAGAAGGATTCGAAGATGT -370

 M T A I D E S V E A Y K K E Y V P M Y E K V L A E K K L L D E E E  29
AATGACAGCAATCGATGAAAGTGTTGAAGCATACAAAAAAGAATATGTTCCAATGTATGAAAAAGTGTTGGCTGAGAAAAAACTATTGGATGAGGAGGAG -360

R K K R D E E                               T L K G K A V H D L N K V  31
AGAAAGAAAAGAGATGAAGAGgtaattgtagtaatttaattctgattatcttcaaattttcagACTCTGAAAGGAAAAGCTGTTCACGACCTGAACAAAG -350

 A N P D E F L E S E L N S K I D R I H L G G V D E E N E E D A E L  35
TCGCCAATCCCGACGAATTTCTGGAGTCGGAGTTGAATTCAAAAATCGATAGAATTCATTTGGGCGGAGTCGATGAAGAGAATGAGGAGGATGCTGAACT -340
          SHP182

 E R S •                                                          35
CGAAAGATCCtgattttcttttgtttttgaatttttattctattttgatccctgtttacttcttattgttctcattttgttgcgttgttttacatttta -330

                    polyA
                    /
ctcatttttgcataaacttgttgcaaaaatcaatataatttttgatctggaaatggttttaaaccttaacctttcatatattaataatttttttcaaaa -320

aaacgttctaaaaaggttcctcatttttcaatataggaaattttgaaga                                                -315
```

Fig 14B

gop-3

SL2

M S E K T F H K 8

tcttttccaaaaatgaggttcttcgcttgaaaagccaacatttaaaacctttttttttccagaaacctagtggttaATGTCTGAAAAGACGTTCCACAAG -3057

A Q T I R A K A S G V P S I V E A V Q F H G V R I T K N D A L V K E 42

GCACAGACCATCCGTGCAAAGGCATCCGGAGTGCCTTCAATCGTCGAAGCTGTACAGTTTCATGGAGTTCGCATCACAAAAAACGATGCTTTGGTTAAGG -2957

V S E L Y R 48

AGgtactacccaaatttcaaaatgttgcacaattcaattgaaaatataaattgtgaattaaattcaacttacatgttttttcagGTTTCCGAATTATACA -285

S K N L D E L V H N S H L A A R H L Q E V G L M D N A V A L I D T 81

GAAGTAAAAATCTAGATGAACTTGTTCATAACTCTCATCTGGCGGCTCGTCATCTTCAAGAAGTTGGATTAATGGATAATGCAGTTGCTCTAATTGATAC -275

SHP183

S P S S N E G Y V V N F L V R E P K S F T A G V K A G V S T N G D 114

ATCTCCAAGCTCAAATGAAGGATATGTTGTCAATTTCCTAGTTCGAGAACCAAAATCATTCACTGCTGGAGTCAAAGCAGGAGTTTCAACGAATGGAGAT -26

A D V S L N A G K Q S V G G R G E A I N T Q Y T Y T V K 14

GCGGATGTCAGTTTAAATGCCGGAAAACAAAGTGTTGGAGGACGAGGAGAGGCAATCAATACACAGTATACATATACTGTAAAGgtaaggacgagagttg -255

SHP145

gcactgccagtttggcatgttctcccaatatttttaattataaaatttggaagtataaaaaaatgtttgcttcatctaaaaatagccttttcacatga -245

aaaaaattgaaaaaaagtgctcaaaaatttcagaaatttccaatttccaaacaattttggagaactttcaaaaatttttccaactgaaattaaagctata -235

Fig 15A

28/32

gop-3 continued...

```
                                                                G D H C F   147
ttctatcactaaattttatacaagtcttaagagaaaatgatgaagtggctcattttgtagaatttcctaaaaaataatatcttcagGGCGATCACTGCTT  -225

N I S A I K P F L G W Q K Y S N V S A T L Y R S L A H M P W N Q S   180
CAACATTTCCGCAATCAAACCATTCCTGGGATGGCAAAAATATTCGAATGTATCAGCGACTCTATACCGTTCACTTGCACATATGCCATGGAATCAATCA  -215
     SHP138                                                                          SHP146

D V D E N A A V L A Y N G Q L W N Q K L L H Q V K L N A             208
GATGTTGATGAGAATGCAGCTGTTCTTGCATATAATGGACAACTATGGAATCAAAAGCTTTTGCATCAAGTCAAATTGAATGCGgtaaagtattataagt  -205

                                I W R T L R A T R D A A F S V R E Q A G H T L   23
gttttgtccaaactatgatacagttcttcagATATGGAGAACACTTCGTGCCACTCGAGATGCCGCATTTTCAGTTCGTGAACAAGCCGGACACACTTTG  -195

K F S L E N A V A V D T R D R P I L A S R G I L A                   25
AAATTCTCGTTGGAGAATGCTGTAGCTGTTGATACAAGAGATAGACCTATTCTTGCAAGTCGTGGAATTCTTGgtaagagtaacaacgactatttttaaa  -185


aaatatctttttcgaaaaaattacgaacgaaaaaaaactgtattatgtacccaaacgcgaaattttgcagttcttgcgcgttcttgttgataaaaaatat  -175

                                                                R F A Q   26
gtaaaaaattggaaaaactacgaaaagtcgataaaaattccgtaccaaccggaaaatgtttcattaatttctcttcctttttttcagCTCGTTTTGCTCAA  -165

E Y A G V F G D A S F V K N T L D L Q                               279
GAGTACGCAGGAGTATTTGGTGATGCGTCATTTGTGAAGAATACATTAGATTTACAGgtaacaaccttatttcaacaattatttcaaattctattaaaaa  -155
       SHP139

      A A A P L P L G F I L A A S F Q A K H L K G L G D R E V H I L   31
taattccagGCAGCTGCCCCTCTTCCACTCGGTTTCATTCTTGCCGCCTCATTCCAAGCGAAACATTTGAAAGGACTCGGAGATCGAGAAGTTCATATTT  -145
        SHP140
```

FIG 15B

gop-3 continued...

```
D R C Y L G G Q Q D V R G F G L N T I G                                  330
TGGATAGATGTTATTTGGGTGGACAACAGGATGTTCGAGGATTTGGTCTGAATACTATTGGAgtgagttttaacgaaattctcttgaaagtcaaataatc -1357
SHP184

V K A D N S C L G G G A S L A G V V H L Y R P L I P P N M L F  361
attttcagGTTAAAGCAGATAACAGTTGTCTTGGAGGAGGTGCTTCACTTGCTGGTGTCGTTCATTTGTATCGGCCATTGATTCCACCAAATATGCTATT -1257

A H A F L A S G S V A S V H S K N L V Q Q L Q D T Q R V S A G F G  394
TGCACACGCATTCCTTGCATCTGGAAGTGTTGCATCAGTTCATTCCAAAAATTTGGTGCAACAATTACAGGATACTCAACGAGTATCAGCCGGATTTGgt -1157
SHP163

gagtttgaaatttaggaaacatttggatgaaatgtatttttaaaaatagatcagctttatttatttgaaaaaaaacgctcattaatcaatagtgatagt -1057

tccattctgagtttcttcttcttcctcgcggaatacaatttttgacttgttcgcatccttcttgtgtactttgtcaccaatcttctcatcaactaaatct -957

cgaaactgaaaaaatttcaaaattattccaaaaaatattgatgcagactacctttttgatggcttctggtacgtttctagcgtcgaatggattggctcct -857

ccaataattaaagtctcgttcggtagtttagccagacggacggtgtgcttcaacatttttctaattaatctatttcaattcaagtcactcactctctctt -757
```

FIG 15C

30/32

gop-3 continued...

```
gacgtcttcttctatattccaagaactctgcagaaaatccgtgtccgccttgtgtgtttctagttggcgtcggaggattcacgggtccaagacgaatgga  -657

tgtctaaaaaatgttatatttttgcataaagaaaacaccataccttcaccacttttgagttgtgggcgttctgaatggaattgatcgattattattgct  -557

ctttcttgatttgcttctatcagctgcgtaatgaggtgttctaaagatcagctttaattcatttggacaagtgctcctctaataaacttaccctgtactc  -457

atttttgaaacgatttacgatgataagattgaaagtggaagttaaatttagtctttcaaagttgaaataaaatcttcataaataaataaatttaaatgaa  -357

                                                                                          L  A  F  V  F  K  S  401
agattaaataaattaacgttcacgtagttaaaaaaataatttaaatcttaaacttctaataaaaaatctcaattttccagGACTCGCATTCGTGTTCAAAA  -257

 I  F  R  L  E  L  N  Y  T  Y  P  L  K  Y  V  L  G  D  S  L  L  G  G  F  H  I  G  A  G  V  N  F  L   434
GTATTTTCCGGCTGGAACTCAACTACACGTATCCATTGAAATATGTGCTCGGCGATTCATTGCTCGGTGGATTCCATATTGGAGCTGGTGTCAACTTCTT  -157

Gtagagattaattggatgcaagcacccctcaaaaagatttttttgaaaaacgataaattcacagaatttcagttctttttctccccctttattgttatt  -57
          SHP134

ttcatcgtaatgctgtgctagaagtcagagtaaatatgagtttttttgtgttctaggaattccatttttttcaggaagcaaatttaataaaaattatcgaa   44
          SHP164
     polyA
tttcttgctctaaagatgttgtacattttatggaaatgttcgtatagtaa                                                      94
                              SHP135
```

FIG 15D

hap-1

SL2

M S L R K I N F V T G 11

ttcgaacactttatatttctcgttttaaaactgtcggtgttttatagtaaactatcttcagaaaaaaATGAGCCTACGAAAAATCAATTTCGTAACTGGA 194

SHP91 SHP118

N V K K L E E V K A I L K N F E 27

AACGTGAAGAAGCTTGAAGAAGTCAAGGCTATTTTGAAGAATTTCGAGgtaaaatatatttgatattattcgaacgcgaaattttgcgccaaaagtacga 294

tgcctggtctcaacacgacaatattttgttaaatacaaacgaatgtgcgccttcaaagaaaagtttcaatctttcgttgccgtggagatatttttagagt 394

V S N V D V D L D E F 38

ttttgtttaaattatatatttgtcgtatcgaaaccgggtaccgtaatcaatcaattaaatattttcagGTTTCAAACGTGGATGTCGATTTGGATGAATT 494

SHP165

Q G E P E F I A E R K C R E A V E A V K G P V L 62

CCAAGGAGAACCCGAATTTATTGCCGAAAGAAAGTGCCGTGAGGCTGTTGAAGCTGTAAAAGGGCCCGTTTTGgtatggaaaattgtatttgttctaaaa 594

V E D T S L C F N A M G G L P G P Y I K W F L K N L K P E 91

attgtcaaatttcagGTCGAAGACACAAGTTTATGCTTCAACGCAATGGGCGGTCTTCCTGGACCTTATATCAAGTGGTTTTTGAAGAATTTGAAACCAG 694

SHP129

FIG 16A

32/32

hap-1 continued...

```
 G L H N M L A                                   G F S D K T A Y A Q C I F   111
AAGGACTACATAATATGCTAGgtaaatattttaatttttttgaaaaaacttatttttcagCCGGATTTTCTGACAAAACCGCCTATGCTCAATGCATCTTT  794

A Y T E G L G K P I H V F A G                                              126
GCGTACACTGAAGGACTCGGAAAACCTATTCATGTATTTGCTGgtatgatttttttgaatttaattctttaattttatatgttaatttagttgtttcattc  894

                                                K C P G Q I V A P R G D T A F G W D P   145
ctcaatttatgagagatttttttttcaatttttctatttcagGAAAATGTCCTGGTCAAATTGTTGCTCCACGTGGTGATACTGCTTTTGGATGGGATCC  994
                                                                    SHP130

 C F Q P D G F K E T F G E M D K D V K N E I S H R A K A L E L L K   178
ATGCTTCCAGCCAGATGGTTTTAAAGAAACATTCGGAGAAATGGATAAAGATGTAAAAAATGAAATTTCTCATCGTGCAAAGGCTCTGGAACTCCTCAAG 1094
                SHP119                                          SHP120

E Y F Q N N •                                                              184
GAATATTTTCAGAATAATtaaattatttttttctcatctatgcaatttcttgaaaatttgttaagtttccgttgttatgcatttgcttttatttaaaaaaa 1194

          polyA
          ⌐
aaagaatatttttacattaatattagatatgagaaaagagtaatttctggattttaaccttcctacaaaagaatatttatatttttttgtatgatttttta 1294
                                SHP93
```

Fig. 16B

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: McGILL UNIVERSITY

(ii) TITLE OF INVENTION: THE C. ELEGANS gro-1 GENE

(iii) NUMBER OF SEQUENCES: 62

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: SWABEY OGILVY RENAULT
(B) STREET: 1981 McGill College Avenue - Suite 1600
(C) CITY: Montréal
(D) STATE: QC
(E) COUNTRY: Canada
(F) ZIP: H3A 2Y3

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Diskette
(B) COMPUTER: IBM Compatible
(C) OPERATING SYSTEM: Windows
(D) SOFTWARE: FastSEQ for Windows Version 2.0b

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: CA 2,210,251
(B) FILING DATE: 25-AUG-1997

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Côté, France
(B) REGISTRATION NUMBER: 4166
(C) REFERENCE/DOCKET NUMBER: 1770-179PCT FC/ld

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 514 845-7126
(B) TELEFAX: 514 288-8389
(C) TELEX:

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14458 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

GCAAAATTTG CTAAGATGAA GCGCCGGCTT GTTACATTGC TTTTCAGAGT CGATTGGTTC 60
AAAATTGTCA ATTTTATCCA AAATAGAGTG CATTGTGTGT ACAATAACTA AAGAATCATC 120
CATATCTGGT CCAACACAAC ATTGATGGAA TACTGGATCA ATTGTCTAAA AAAATATCAA 180

```
TAGAATAATG AAACATTTTC AGAATTCATT ACCGTCAATG TCAGATAGTC ATTCCTTGAG    240
TATTTTGTGG ATGCTTTGAA AATTCTTCGC TGGGCCATAT CTGTTGGATA ATCTGAAAAA    300
CGCAATAAAT TTCATCGAAA ATGCCTATTA AATTGAATTA CCTTCTTCTT CATCATTTCC    360
TAACAATTCA TGCTCTTTTT GTGCTTGACT TGTGACCAAT TCTTTAAATT CAATTAAATC    420
GTCAATATCC TTTTGTACTA AATCCATCTT GATATTCAAT ATATCTTTGT CAGTATAGTA    480
TTCAGCGTAT CTGAAATTTC GAATTTATTT TTCTAATTCC CAAGAAAAAT AATTAATAAG    540
AATACCTTAA CGAATTATTA TCCAATATAT CATCATTTGC CACATCTGGA AGACGCTGAG    600
GAACTGTTTG AGCAGCTTGG AGGTAGTCGT CATCGTCTCT GGAAATTGTT ATTTTCAATT    660
TCAAAAAAAA AACTTTACTT ACGAAATATA CTCATTTGAT GCAATCCACG GATCAAAACG    720
ACGTCTTTGC ATCTTTGAAT CATTTTCCGC ATGGCACCGC ATCACTTCTT TCTTATGATT    780
ATTTTCTAAC GTTTTTGAAA ATTCGACGTG CTCTTCACAA CGGCCGCCAT GTTTCGCAAG    840
TTCTTCTTTT GATCGTATCT AAAATTTTAA ATTTGAAAAA AAGCTTACTA TCAAATTTTC    900
GTATTTTTTC TCACCTGCTT ACACCGAACA AGCGTTCGAT ACGAAGCATA ATTACATTGT    960
CCATACTTAT TTTTGTCGTA TTCATTGGCA ACAAGACGGA ATCGTGTTCC AGGTGCAACT   1020
ATATATTGAG CAGGAGGACG AGTTGTTTGT TTCATGCTGC TTAAAAATAA AAATGGAAAA   1080
TTGAGTCAAA AAGTTGAGAT AAAACAAATT AAAACAATTT TCTGAAAAAT AAACAACTGA   1140
AATTTGAAGT AATAAACAAC ACGCGAAAAC GTTATTTCGG AGCATCGTTT GAGAAGTAAA   1200
ACTTTTTTTC GGCGCACCCT TGTGCGCAGT TTTTATCTTC TCTTTTAATT TAATTTTCAA   1260
GCTAAATCTT TCTTTTTAAA CTTTGAATAA ATATTTAAAT ATTCAGAATG CACCAATAAA   1320
CCTGGAACAA AATCGATAAT GTTCCGCAAG CTTGGTTCTT CTGGGTCACT ATGGAAGCCG   1380
AAAAATCCGC ATTCTTTGGA ATACCTCAAA TATTTACAAG GAGTGCTCAC AAAAAATGAG   1440
AAAGTTACGG AAAACAATAA GAAAATATTA GTAGAAGCAT TACGAGCTAT CGCAGAAATT   1500
CTCATTTGGG GCGATCAGAA TGATGCTTCG GTTTTTGAGT GAGTTTTTTT CCAATGTTTT   1560
TTTTCAAATC TGATGTTGAA TTTCAGTTTC TTCCTTGAGC GGCAAATGCT TCTTTATTTC   1620
TTGAAAATTA TGGAACAAGG AAACACACCA CTAAATGTAC AATTACTGCA GACTTTGAAC   1680
ATTTTATTCG AAAATATTCG ACATGAAACT TCACTTTGTA AGTTTTTTAT ATGGATTTTC   1740
GCTTAAAATT GCCAGTTTTC AGATTTCCTT CTAAGTAACA ATCATGTAAA CTCGATTATT   1800
TCCCACAAAT TCGATTTACA AAATGATGAG ATCATGGCTT ACTACATTAG TTTTCTGAAA   1860
ACTCTTTCAT TTAAACTGAA TCCAGCTACA ATCCACTTCT TCTTCAATGA AACGACTGAA   1920
GAATTTCCAT TGTTGGTAGA AGTTTTGAAG CTTTATAATT GGAATGAATC AATGGTTCGA   1980
ATTGCTGTTA GAAATATTCT TTTAAATATT GTGAGAGTTC AAGATGATTC AATGATTATT   2040
TTCGCTATCA AGCATACAAA AGTTAGTAGA AAATTATTTT GAAAAGGTGT ATTTAAGCAA   2100
TAAATATTAC AGGAATATCT ATCGGAGTTA ATAGATTCTC TAGTTGGTCT CTCACTTGAA   2160
ATGGACACAT TTGTACGATC TGCTGAGAAT GTGTTAGCTA ATCGAGAGAG ATTACGAGGA   2220
AAAGTGGATG ATTTAATTGA TTTGATTCAT TATATTGGTG AACTATTGGA TGTGGAAGCT   2280
GTCGCCGAAA GTTTATCAAT TTTAGGTCAG TTTTACTGCT GGAAAATCAA GTTTTTAATG   2340
TTAAATTTTC AGTAACAACA CGATACTTAA GCCCTCTATT ACTTTCAAGT ATATCACCAA   2400
GAAGAGATAA TCATTCACTT CTACTCACTC CGATTTCTGC GTTATTTTTT TTCTCTGAAT   2460
TTTTATTGGT GAGTTTTAAC ATTTAAAATT ACATTTTTCT AATTTATTTA TTTTTCAGAT   2520
AGTTCGTCAC CATGAAACAA TATATACATT TTTATCATCT TTCCTATTTG ACACTCAGAA   2580
TACTTTGACG ACCCATTGGA TACGTCATAA TGAGAAATAT TGCTTAGAAC CGATTACATT   2640
ATCATCACCA ACCGGAGAAT ATGTGAATGA AGACCAGTAA GAGCTGAAAT TTTAAAATTT   2700
TTGCTTTGAA TATAGTATTT TCAGCGTATT TTTCGATTTT CTACTGGAAG CATTTGATTC   2760
CAGTCAAGCA GACGATTCGA AGGCATTCTA TGGATTAATG CTGATTTATT CAATGTTTCA   2820
GAATAATGGT GAGTTTTAAA AAATTGATTT GTTAAATTAA AATTTCCATT TCCAATAACT   2880
CCTCTTCAGA CAGTAAGTTT TCAATGTTGT AAAGTTCCTG TTCATCTGTG ATCGTTTTCT   2940
TCATTTTTTT AGTTTTGCAT GAACAGTTTT CAAATTTTTT TGATATCATA CAGTAAATAT   3000
CGTCATCCAG ATAATTTTCT ATTTAAAAAA AATGAATAAA AAGAGGGCGC GCAGAAATTG   3060
CCGAAGTAAT GTAAATTTAA AGGGACACAT GCGTAGCTTG TTGTGTGGGT CTCGCCGCGC   3120
TTTGTTTGAT TTATCTTGTT TTCTGCTCAA AGAGCTGTTT TTATTTTAGC GTTGAATGCT   3180
TTTTTACCGT TCTCATCGGC TTTTTAATAG GAATATTTAA AAAAAAGGT TTAATAAATC   3240
TTCGTTTTTA CAAAATCCAT CTAAGATTTG CATTTGTGAA GCTCAACAAG TAAAGTTTTA   3300
AGTAACATTG TTTTTTAAAA AACAATTGAA CCAAATTTTG CCGAAACATT AATAACATGA   3360
CGATACTCTA TAAAATATTC CTCTTTTCAA AATAAATTTT CAAAAAAAAT CCATTTTTCA   3420
GCCGATGTTG GAGAACTTCT ATCTGCTGCC AACTTCCCAG TGCTCAAAGA ATCAACGACA   3480
ACTTCATTAG CTCAACAGAA TCTTGCTCGT CTCCGAATAG CATCTACGTC TTCCATATCA   3540
AAGCGAACGA GAGCTATCAC TGAAATTGGA GTAGAAGCGA CCGAGGAAGA TGAGATTTTT   3600
CATGATGTTC CTGAAGAACA AACGTTGGTA AGTAAATAAA TCAACATTGA TTGTTACACA   3660
AACTTTAATA TTTTTAAATT TGAAAATTTT CTTCAAAGTG CTCAAAAATC CTGTCGAAAA   3720
```

```
TTACAGGAAG ATCTGGTGGA TGATGTATTG GTTGATACTG AAAATTCAGC AATAAGTGAT   3780
CCAGAAGTGA GTAGAAAACG TGCATGTATT AATTATTAAA AAAAAAATAT AGTTTTCCCC   3840
AGTTTTCCTT GACCTAAAAC TCAGCAATTT CAGCCTAAAA ACGTGGAGTC AGAATCTCGT   3900
TCTCGATTTC AATCTGCTGT TGATGAGCTT CCACCTCCGT CGACTTCTGG ATGTGATGGT   3960
CGACTTTTTG ATGCACTTTC ATCGATTATC AAAGCAGTTG GAACAGATGA CAATCGAATT   4020
CGACCAATTA CATTGGAACT TGCATGTCTT GTAATTCGGC AAATTTTAAT GACTGTTGAT   4080
GATGAAAAAG TAAGATTACA AATTCAAAAT TGAGCAAAAT CAGAATCTAA ATTTCATAAA   4140
TTGTTCAGGT ACATACCAGT TTAACGAAAT TATGCTTCGA AGTTCGTCTA AAACTTTTAT   4200
CATCAATTGG ACAATATGTT AATGGAGAGA ATCTGTTTTT GGAGTGGTTT GAGGATGAAT   4260
ATGCAGAATT TGAAGTAAGC CAAGAGGTCC GAAAATAATT TAATTCATCC TTTTTATTCA   4320
GGTGAATCAC GTGAATTTCG ATATAATCGG TCACGAAATG CTTCTTCCTC CAGCTGCAAC   4380
TCCTCTTTCG AATCTGCTAC TTCATAAGCG ATTGCCCAGT GGATTTGAAG AACGAATAAG   4440
AACTGTAGGA AACTTTTTAA ATTTGAAAAT TAATTATATA TATATTTGCA GCAAATCGTA   4500
TTCTACCTAC ATATTCGAAA ATTGGAACGA GATTTGACCG GTGAAGGAGA CACAGAATTA   4560
CCTGTGAGAG TGTTGAATTC TGATCAGGAA CCAGTTGCCA TCGGTGATTG TATTAATTTA   4620
CGTGAGTTCA TCTGCATAGA AAACACCATA TTTCTACTCA AATTAACAAT TTTCAGATAA   4680
TTCGGATCTT CTATCCTGCA CTGTGGTTCC TCAACAACTA TGTTCTCTTG GAAAACCTGG   4740
TGATCGTCTT GCTCGATTCC TTGTCACTGA TAGACTTCAA TTAATTCTTG TCGAACCGGA   4800
TTCTCGAAAA GCCGGATGGG CAATTGTTCG ATTCGTAGGA CTTCTTCAAG ATACAACAAT   4860
TAATGGAGAT TCTACGGATT CGAAAGTTTT GCATGTTGTG GTGGAAGGGC AACCCTCGAG   4920
AATTAAGGTA AGAATACTAA CGGGAAAAAA AAATCAAAAA ATTACTTCTG TTTCAGAAAA   4980
GACATCCGGT TTTAACTGCA AAGTTCATAT TCGATGATCA CATTCGGTGT ATGGCAGCAA   5040
AGCAACGGCT CACCAAGGTA ACGGAAAAAA TAACCAAAAA GACGGAAAGT TATTGTAAAT   5100
GGACGAAATC GGCGAAATTA ATTGAAAACG TTTGAATTTG CCGCTAAAAC CAAACGAAAA   5160
CCAAACGAAA GCGAAATTTA ACTATCCCTT CAGGTAGAAT ATACATTTTA TTTCTCTTTA   5220
TAGGGTCGCC AAACAGCACG TGGTCTGAAA CTTCAGGCGA TATGTTCAGC TCTTGGAGTT   5280
CCACGTATCG ATCCAGCGAC AATGACGTCA TCACCACGAA TGAATCCATT CAGAATTGTG   5340
AAAGGATGCG CACCGGGAAG TGTACGAAAA ACTGTTTCCA CATCATCATC GTCAAGCCAA   5400
GGACGTCCCG GACATTATTC TGCAAATCTT AGATCAGCAT CTAGAAATGC AGGAATGATA   5460
CCAGATGATC CAACTCAACC GAGTAGTTCT TCGGAAAGAA GATCCTAGGG ATCAATATCT   5520
CTTCAGTTTC ATCATTTTAT GCTGTAAATT GTATTTAAGT ATTCCTATTC TTTGTAGTAC   5580
TGTATTTACA CATCGTCTAG TTAAAATCAC AAATCTCCGA AAAAACAAAC CAGTGAACAT   5640
GTGATATTTC TCTTGCCCAT AGTTCTCTTT TTTTTTTGAA ACAAAAACAA TTACTTTTAT   5700
GCTCACCTAT TCGAGCCATA TTTTTTTCCC AATTACCGGT TGTTTATTTT AATTTCTTTT   5760
TTTTTTCTGT AAATCTACTT TATTTTTAAA ACTGCATTTG AGATTGTGTA TATTTTTTCA   5820
AAATGGTTCA AATGCCGAAT CTATCTACTT TTTAATCATT ATTCAAACAG AAAAACCGAT   5880
TATTTATTCA GATTCTCAAA AATGGCTGAA AAAGCTGAAA ATCTTCCATC TTCTTCGGCC   5940
GAAGCTTCAG AAGAGCCATC ACCTCAAACT GGACCAAATG TGAATCAAAA ACCATCGATT   6000
TTGGTTCTTG GAATGGCTGG TTCTGGAAAA ACGACATTTG TTCAGGTAAC TTTCATTCAA   6060
TTTTGAGAGT TTTCAAACAT TACTATTTTC AGCGTCTCAC AGCATTCCTA CATGCTCGTA   6120
AAACACCTCC ATATGTGATT AATCTGGATC CGGCAGTTAG CAAAGTACCT TATCCAGTGA   6180
ATGTTGACAT TCGAGATACT GTGAAATACA AGGAAGTTAT GAAAGAATTC GGAATGGGAC   6240
CAAATGGAGC AATTATGACA TGTCTTAACC TGATGTGTAC TCGTTTTGAT AAAGTAATTG   6300
AGTTGATTAA TAAGAGATCT TCTGATTTCT CAGTTTGTCT TCTTGATACT CCTGGACAAA   6360
TTGAAGCATT CACTTGGAGT GCTAGTGGAT CTATTATCAC TGATTCATTG GCAAGTAGCC   6420
ATCCCACGGT AAGGGATTTT GATTTATGAA ATCTGCTTGA AATGAAAAAA GATTCTAATA   6480
AATTTTTGAC TTTTAAACAT TTTTTACAGT TATATTTGGT CTATTTTCTA TCATTAAAAG   6540
CAAAATGAAA AGTCGATTCT ACTCCATATT TATTAATTTC GACTTTTCAG GTGGTAATGT   6600
ACATTGTGGA TTCCGCTCGT GCCACAAATC CAACTACATT CATGTCCAAT ATGCTCTACG   6660
CATGTTCCAT TCTCTACCGT ACCAAACTTC CATTCATTGT CGTTTTCAAC AAAGCTGATA   6720
TTGTCAAACC AACATTTGCA CTCAAATGGA TGCAAGATTT CGAAAGATTT GATGAAGCTT   6780
TAGAGGATGC CAGAAGCAGT TATATGAATG ATTTGAGTCG TTCATTGAGT CTCGTTCTTG   6840
ATGAATTCTA TTGCGGACTG AAAACAGGTT TTTATTCGAA ATAAAACCTT TTTTAAATAA   6900
TAAATTTCAG TTTGCGTCAG TTCTGCAACT GGAGAAGGAT TCGAAGATGT AATGACAGCA   6960
ATCGATGAAA GTGTTGAAGC ATACAAAAAA GAATATGTTC CAATGTATGA AAAAGTGTTG   7020
GCTGAGAAAA AACTATTGGA TGAGGAGGAG AGAAAGAAAA GAGATGAAGA GGTAATTGTA   7080
GTAATTTAAT TCTGATTATC TTCAAATTTT CAGACTCTGA AAGGAAAAGC TGTTCACGAC   7140
CTGAACAAAG TCGCCAATCC CGACGAATTT CTGGAGTCGG AGTTGAATTC AAAAATCGAT   7200
AGAATTCATT TGGGCGGAGT CGATGAAGAG AATGAGGAGG ATGCTGAACT CGAAAGATCC   7260
```

```
TGATTTTCTT TTTGTTTTTG AATTTTTATT CTATTTTGAT CCCTGTTTAC TTCTTATTGT   7320
TCTCATTTTG TTGCGTTGTT TTACATTTTA CTCATTTTTG CATAAACTTG TTGCAAAAAT   7380
CAATATAATT TTTGATCTGG AAATGGTTTT AAACCTTAAC CTTTCATATA TTAATAATTT   7440
TTTTTCAAAA AAACGTTCTA AAAAGGTTCC TCATTTTTTC AATATAGGAA ATTTTGAAGA   7500
TCTTTTCCAA AAATGAGGTT CTTCGCTTGA AAAGCCAACA TTTAAAACCT TTTTTTTTCC   7560
AGAAACCTAG TGGTTAATGT CTGAAAAGAC GTTCCACAAG GCACAGACCA TCCGTGCAAA   7620
GGCATCCGGA GTGCCTTCAA TCGTCGAAGC TGTACAGTTT CATGGAGTTC GCATCACAAA   7680
AAACGATGCT TTGGTTAAGG AGGTACTACC CAAATTTCAA AATGTTGCAC AATTCAATTG   7740
AAAATATAAA TTGTGAATTA AATTCAACTT ACATGTTTTT TCAGGTTTCC GAATTATACA   7800
GAAGTAAAAA TCTAGATGAA CTTGTTCATA ACTCTCATCT GGCGGCTCGT CATCTTCAAG   7860
AAGTTGGATT AATGGATAAT GCAGTTGCTC TAATTGATAC ATCTCCAAGC TCAAATGAAG   7920
GATATGTTGT CAATTTCCTA GTTCGAGAAC CAAAATCATT CACTGCTGGA GTCAAAGCAG   7980
GAGTTTCAAC GAATGGAGAT GCGGATGTCA GTTTAAATGC CGGAAAACAA AGTGTTGGAG   8040
GACGAGGAGA GGCAATCAAT ACACAGTATA CATATACTGT AAAGGTAAGG ACGAGAGTTG   8100
GCACTGCCAG TTTGGCATGT TCTCCCAATA TTTTTTAATT ATAAAATTTG GAAGTATAAA   8160
AAAATGTTTG CTTCATCTAA AAATAGCCTT TTTCACATGA AAAAAATTGA AAAAAAGTGC   8220
TCAAAAATTT CAGAAATTTC CAATTTCCAA ACAATTTTGG AGAACTTTCA AAAATTTTTC   8280
CAACTGAAAT TAAAGCTATA TTCTATCACT AAATTTTATA CAAGTCTTAA GAGAAAATGA   8340
TGAAGTGGCT CATTTTGTAG AATTTCCTAA AAAATAATAT CTTCAGGGCG ATCACTGCTT   8400
CAACATTTCC GCAATCAAAC CATTCCTGGG ATGGCAAAAA TATTCGAATG TATCAGCGAC   8460
TCTATACCGT TCACTTGCAC ATATGCCATG GAATCAATCA GATGTTGATG AGAATGCAGC   8520
TGTTCTTGCA TATAATGGAC AACTATGGAA TCAAAAGCTT TTGCATCAAG TCAAATTGAA   8580
TGCGGTAAAG TATTATAAGT GTTTTGTCCA AACTATGATA CAGTTCTTCA GATATGGAGA   8640
ACACTTCGTG CCACTCGAGA TGCCGCATTT TCAGTTCGTG AACAAGCCGG ACACACTTTG   8700
AAATTCTCGT TGGAGAATGC TGTAGCTGTT GATACAAGAG ATAGACCTAT TCTTGCAAGT   8760
CGTGGAATTC TTGGTAAGAG TAACAACGAC TATTTTTAAA AAATATCTTT TTCGAAAAAA   8820
TTACGAACGA AAAAAAACTG TATTATGTAC CCAAACGCGA AATTTTGCAG TTCTTGCGCG   8880
TTCTTGTTGA TAAAAAATAT GTAAAAAATT GGAAAAACTA CGAAAAGTCG ATAAAAATTC   8940
CGTACCAACC GGAAAATGTT TCATTAATTT CTCTTCCTTT TTTCAGCTCG TTTTGCTCAA   9000
GAGTACGCAG GAGTATTTGG TGATGCGTCA TTTGTGAAGA ATACATTAGA TTTACAGGTA   9060
ACAACCTTAT TTCAACAATT ATTTCAAATT CTATTAAAAA TAATTCCAGG CAGCTGCCCC   9120
TCTTCCACTC GGTTTCATTC TTGCCGCCTC ATTCCAAGCG AAACATTTGA AAGGACTCGG   9180
AGATCGAGAA GTTCATATTT TGGATAGATG TTATTTGGGT GGACAACAGG ATGTTCGAGG   9240
ATTTGGTCTG AATACTATTG GAGTGAGTTT TAACGAAATT CTCTTGAAAG TCAAATAATC   9300
ATTTTCAGGT TAAAGCAGAT AACAGTTGTC TTGGAGGAGG TGCTTCACTT GCTGGTGTCG   9360
TTCATTTGTA TCGGCCATTG ATTCCACCAA ATATGCTATT TGCACACGCA TTCCTTGCAT   9420
CTGGAAGTGT TGCATCAGTT CATTCCAAAA ATTTGGTGCA ACAATTACAG GATACTCAAC   9480
GAGTATCAGC CGGATTTGGT GAGTTTGAAA TTTAGGAAAC ATTTGGATGA AATGTATTTT   9540
TTAAAAATAG ATCAGCTTTA TTTATTTGAA AAAAAACGCT CATTAATCAA TAGTGATAGT   9600
TCCATTCTGA GTTTCTTCTT CTTCCTCGCG GAATACAATT TTTGACTTGT TCGCATCCTT   9660
CTTGTGTACT TTGTCACCAA TCTTCTCATC AACTAAATCT CGAAACTGAA AAAATTTCAA   9720
AATTATTCCA AAAAATATTG ATGCAGACTA CCTTTTTGAT GGCTTCTGGT ACGTTTCTAG   9780
CGTCGAATGG ATTGGCTCCT CCAATAATTA AAGTCTCGTT CGGTAGTTTA GCCAGACGGA   9840
CGGTGTGCTT CAACATTTTT CTAATTAATC TATTTCAATT CAAGTCACTC ACTCTCTCTT   9900
GACGTCTTCT TCTATATTCC AAGAACTCTG CAGAAAATCC GTGTCCGCCT TGTGTGTTTC   9960
TAGTTGGCGT CGGAGGATTC ACGGGTCCAA GACGAATGGA TGTCTAAAAA ATGTTATATT  10020
TTTGCATAAA GAAAACACCA TACCTTCACC ACTTTTTGAG TTGTGGGCGT TCTGAATGGA  10080
ATTGATCGAT TATTATTGCT CTTTCTTGAT TTGCTTCTAT CAGCTGCGTA ATGAGGTGTT  10140
CTAAAGATCA GCTTTAATTC ATTTGGACAA GTGCTCCTCT AATAAACTTA CCCTGTACTC  10200
ATTTTTGAAA CGATTTACGA TGATAAGATT GAAAGTGGAA GTTAAATTTA GTCTTTCAAA  10260
GTTGAAATAA AATCTTCATA AATAAATAAA TTTAAATGAA AGATTAAATA AATTAACGTT  10320
CACGTAGTTA AAAAAATAAT TTAAATCTTA ACTTCTAATA AAAAATCTCA ATTTTCCAGG  10380
ACTCGCATTC GTGTTCAAAA GTATTTTCCG GCTGGAACTC AACTACACGT ATCCATTGAA  10440
ATATGTGCTC GGCGATTCAT TGCTCGGTGG ATTCCATATT GGAGCTGGTG TCAACTTCTT  10500
GTAGAGATTA ATTGGATGCA AGCACCCCTC AAAAAGATTT TTTTGAAAAA CGATAAATTC  10560
ACAGAATTTC AGTTCTTTTT CTCCCCCTTT TATTGTTATT TTCATCGTAA TGCTGTGCTA  10620
GAAGTCAGAG TAAATATGAG TTTTTTTGTG TTCTAGGAAT TCCATTTTTT CAGGAAGCAA  10680
ATTTAATAAA AATTATCGAA TTTCTTGCTC TAAAGATGTT GTACATTTTA TGGAAATGTT  10740
CGTATAGTAA TTCGAACACT TTATATTTCT CGTTTTAAAA CTGTCGGTGT TTTATAGTAA  10800
```

```
ACTATCTTCA GAAAAAAATG AGCCTACGAA AAATCAATTT CGTAACTGGA AACGTGAAGA  10860
AGCTTGAAGA AGTCAAGGCT ATTTTGAAGA ATTTCGAGGT AAAATATATT TGATATTATT  10920
CGAACGCGAA ATTTTGCGCC AAAAGTACGA TGCCTGGTCT CAACACGACA ATATTTTGTT  10980
AAATACAAAC GAATGTGCGC CTTCAAAGAA AAGTTTCAAT CTTTCGTTGC CGTGGAGATA  11040
TTTTTAGAGT TTTTGTTTAA ATTATATATT TGTCGTATCG AAACCGGGTA CCGTAATCAA  11100
TCAATTAAAT ATTTTCAGGT TTCAAACGTG GATGTCGATT TGGATGAATT CCAAGGAGAA  11160
CCCGAATTTA TTGCCGAAAG AAAGTGCCGT GAGGCTGTTG AAGCTGTAAA AGGGCCCGTT  11220
TTGGTATGGA AAATTGTATT TGTTCTAAAA ATTGTCAAAT TTCAGGTCGA AGACACAAGT  11280
TTATGCTTCA ACGCAATGGG CGGTCTTCCT GGACCTTATA TCAAGTGGTT TTTGAAGAAT  11340
TTGAAACCAG AAGGACTACA TAATATGCTA GGTAAATATT TTAATTTTTT GAAAAAACTT  11400
ATTTTTCAGC CGGATTTTCT GACAAAACCG CCTATGCTCA ATGCATCTTT GCGTACACTG  11460
AAGGACTCGG AAAACCTATT CATGTATTTG CTGGTATGAT TTTTTGAATT TAATTCTTTA  11520
ATTTTATATG TTAATTTAGT TGTTTCATTC CTCAATTTAT GAGAGATTTT TTTTTCAATT  11580
TTTCTATTTC AGGAAAAATGT CCTGGTCAAA TTGTTGCTCC ACGTGGTGAT ACTGCTTTTG  11640
GATGGGATCC ATGCTTCCAG CCAGATGGTT TTAAAGAAAC ATTCGGAGAA ATGGATAAAG  11700
ATGTAAAAAA TGAAATTTCT CATCGTGCAA AGGCTCTGGA ACTCCTCAAG GAATATTTTC  11760
AGAATAATTA AATTATTTTT TCTCATCTAT GCAATTTCTT GAAAATTTGT TAAGTTTCCG  11820
TTGTTATGCA TTTGCTTTTA TTTAAAAAAA AAAGAATATT TTTACATTAA TATTAGATAT  11880
GAGAAAAGAG TAATTTCTGG ATTTTAACCT TCCTACAAAA GAATATTTAT ATTTTTTGTA  11940
TGATTTTTTA AAAATATCGT CAGGAAATAA TAACATTTCA GATATACCCT GAACTCTACA  12000
GTTTATGATA TTCAGGAAAT TTCTGAATTT TCTGAAACCT TACAAAATGC GAACGGATCC  12060
GATTATTTTC GTGATTGGGT GCACTGGAAC CGGGAAAAGT GATCTTGGAG TGGCAATTGC  12120
AAAGAAATAT GGAGGAGAGG TGATTAGTGT AGATTCAATG CAATTTTATA AAGGTACATG  12180
GGTTTTGTTT CAATTTTAAA TTAATTAATT TTCGTTTTTC AGGACTTGAC ATTGCCACGA  12240
ATAAGATAAC GGAAGAAGAA TCTGAAGGGA TTCAACATCA TATGATGTCA TTTTTGAATC  12300
CATCTGAATC ATCATCTTAT AATGTACATA GTTTCCGAGA AGTCACGTTG GATCTTATTA  12360
AAGTGCTTAA TTCGCCACTT TTTGAACTTG ATCCTAATTT TCATAATTTT CAGAAAATCC  12420
GCGCCCGTTC AAAAATTCCT GTAATTGTCG GAGGAACCAC TTATTATGCT GAAAGTGTCC  12480
TTTATGAGAA TAATCTGATT GAAACCAACA CTTCAGATGA CGTGGATTCC AAATCGAGAA  12540
CATCATCAGA ATCGTCATCT GAAGACACTG AAGAAGGAAT TAGTAATCAA GAATTATGGG  12600
ATGAATTGAA AAAAATCGAC GAAAAATCAG CACTTCTTCT ACATCCAAAT AATCGTTATC  12660
GAGTACAGAG AGCATTGCAA ATTTTCAGAG AAACTGGTAA TTGATTTGCA AATTTCCAGA  12720
TTAAAAACAA ATCAAGTAAA GTTTTTTGCA GGAATCCGAA AAAGTGAACT TGTTGAAAAA  12780
CAGAAATCAG ATGAAACTGT TGATTTGGGT GGACGACTAC GATTTGATAA TTCTTTAGTT  12840
ATTTTTATGG ATGCAACACC TGAAGTTTTA GAAGAAAGAC TTGATGGAAG AGTTGATAAA  12900
ATGATTAAAT TGGGTTTGAA GAATGAATTG ATCGAGTTTT ATAACGAGGT AAATATTTGA  12960
ATTTTTCCAG AAAAAAAAAG AAAATTTTTT ATTATTTTGT TTTTTTTTCA TTCTTTACTA  13020
TTTTCCAAAA AAGTTTAAAC TTTTGAAAAC TGTTCAGAAA ATGTTCGTGT ATTTATTTTA  13080
GCTTACTGAG GCATTATTTC ATTGTGATTT TTACTATACT CTATAAACTA AATTTTCAGC  13140
ACGCCGAGTA CATAAATCAC AGCAAATATG GTGTCATGCA ATGTATTGGT CTTAAAGAAT  13200
TCGTTCCATG GCTCAATTTG GACCCATCAG AAAGAGATAC ACTCAATGGG GATAAATTGT  13260
TCAAGCAAGG GTAATTTAAA TTTATTTTCA ATTTTTATAA ATTCCAAGCT ATTTTCAGAT  13320
GCGATGATGT GAAGCTTCAC ACTCGACAAT ATGCACGGCG CCAGAGACGG TGGTATCGAT  13380
CGAGACTTTT AAAACGGTCG GATGGTGATC GGGTATGTTG ATTTTAAAAA AATTGAATTT  13440
TTAAAGAACT TTTTTACTAA ATTAACAAAG TTATTGGCTG AAAATGGCTG AAAATTATAG  13500
TAAAACTAAT CAAAAAAATT GAAATTTTGA ATTAAAGTCA TAAAGTGACG ACCAGAAAAT  13560
TAAAAAAAAA CATTTTTCTA TTTTAATTAA TTCACTCTAC TTCACTTTAA AAATAATTTT  13620
CAGAAAATGG CAAGTACAAA AATGCTGGAT ACATCTGACA AGTACCGAAT AATTAGTGAT  13680
GGAATGGACA TTGTTGATCA ATGGATGAAT GGAATCGATC TATTTGAAGA TGTAAAATTT  13740
CACAAATTCT AAAATTTCCG AATCACAAAT TAAAATTTCT ACAGATCTCC ACAGACACCA  13800
ATCCAATTCT AAAAGGGTCC GATGCAAATA TTCTGCTGAA TTGTGAAATC TGTAATATTT  13860
CAATGACTGG AAAAGATAAT TGGTTTGTTT CAATACATAT TATAATTTCG AAATGAATTT  13920
TTTCAGGCAG AAACATATCG ATGGGAAAAA GCACAAGCAT CATGCTAAGC AAAAGAAATT  13980
GGCAGAGACT CGCACATAAG ACGCTATATT TATTTTTTGT TAACTTAAAT TATTTTTGTT  14040
GTTGATTGTT CTCTAAATAA AAAAACAGCT CAGAGAGAAG ATTAGGCGCT CGTCCACATC  14100
TCCGACGATA GTCAACCCGA ACGAAGGGAA CTATCTTTAA TTGTCAGTGA TGACGTCATG  14160
TCGTCAAGAA CTCGTCATAG CTGTGAGAAT TGAACCATTA TAGATTTGGA CATTAGTTTA  14220
GGTTATATCC AGTACACTAA ATGGTACATG ATAGACAGTG TACATTTACA GATTTATAGA  14280
TTGTCTCAGT GACTAGTTAC CGGAAGAGGA GAGGAGAACA TGTGGCGATG TCTTTTGGAT  14340
```

CGATATTATT CCGTCTGAAA ATTGTTCACT AGGGGGACTG CCGATTACCA CTTCACATGA 14400
CGGAACATGT TAGTTAAAAT ATTGGCTTTT ATACACATTT TCAAAATAGC ACCTGTAT 14458

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 430 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
Met Ile Phe Arg Lys Phe Leu Asn Phe Leu Lys Pro Tyr Lys Met Arg
 1               5                  10                  15
Thr Asp Pro Ile Ile Phe Val Ile Gly Cys Thr Gly Thr Gly Lys Ser
            20                  25                  30
Asp Leu Gly Val Ala Ile Ala Lys Lys Tyr Gly Gly Glu Val Ile Ser
        35                  40                  45
Val Asp Ser Met Gln Phe Tyr Lys Gly Leu Asp Ile Ala Thr Asn Lys
    50                  55                  60
Ile Thr Glu Glu Glu Ser Glu Gly Ile Gln His His Met Met Ser Phe
65                  70                  75                  80
Leu Asn Pro Ser Glu Ser Ser Ser Tyr Asn Val His Ser Phe Arg Glu
                85                  90                  95
Val Thr Leu Asp Leu Ile Lys Lys Ile Arg Ala Arg Ser Lys Ile Pro
            100                 105                 110
Val Ile Val Gly Gly Thr Thr Tyr Tyr Ala Glu Ser Val Leu Tyr Glu
        115                 120                 125
Asn Asn Leu Ile Glu Thr Asn Thr Ser Asp Asp Val Asp Ser Lys Ser
    130                 135                 140
Arg Thr Ser Ser Glu Ser Ser Ser Glu Asp Thr Glu Glu Gly Ile Ser
145                 150                 155                 160
Asn Gln Glu Leu Trp Asp Glu Leu Lys Lys Ile Asp Glu Lys Ser Ala
                165                 170                 175
Leu Leu Leu His Pro Asn Asn Arg Tyr Arg Val Gln Arg Ala Leu Gln
            180                 185                 190
Ile Phe Arg Glu Thr Gly Ile Arg Lys Ser Glu Leu Val Glu Lys Gln
        195                 200                 205
Lys Ser Asp Glu Thr Val Asp Leu Gly Gly Arg Leu Arg Phe Asp Asn
    210                 215                 220
Ser Leu Val Ile Phe Met Asp Ala Thr Pro Glu Val Leu Glu Glu Arg
225                 230                 235                 240
Leu Asp Gly Arg Val Asp Lys Met Ile Lys Leu Gly Leu Lys Asn Glu
                245                 250                 255
Leu Ile Glu Phe Tyr Asn Glu His Ala Glu Tyr Ile Asn His Ser Lys
            260                 265                 270
Tyr Gly Val Met Gln Cys Ile Gly Leu Lys Glu Phe Val Pro Trp Leu
        275                 280                 285
Asn Leu Asp Pro Ser Glu Arg Asp Thr Leu Asn Gly Asp Lys Leu Phe
    290                 295                 300
Lys Gln Gly Cys Asp Asp Val Lys Leu His Thr Arg Gln Tyr Ala Arg
305                 310                 315                 320
Arg Gln Arg Arg Trp Tyr Arg Ser Arg Leu Leu Lys Arg Ser Asp Gly
                325                 330                 335
```

```
Asp Arg Lys Met Ala Ser Thr Lys Met Leu Asp Thr Ser Asp Lys Tyr
            340                 345                 350
Arg Ile Ile Ser Asp Gly Met Asp Ile Val Asp Gln Trp Met Asn Gly
        355                 360                 365
Ile Asp Leu Phe Glu Asp Ile Ser Thr Asp Thr Asn Pro Ile Leu Lys
    370                 375                 380
Gly Ser Asp Ala Asn Ile Leu Leu Asn Cys Glu Ile Cys Asn Ile Ser
385                 390                 395                 400
Met Thr Gly Lys Asp Asn Trp Gln Lys His Ile Asp Gly Lys Lys His
                405                 410                 415
Lys His His Ala Lys Gln Lys Lys Leu Ala Glu Thr Arg Thr
            420                 425                 430
```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2041 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
CTGCCATAAG ATGGCGTCCG TGGCGGCTGC ACGAGCAGTT CCTGTGGGCA GTGGGCTCAG     60
GGGCCTGCAA CGGACCCTAC CTCTTGTAGT GATTCTCGGG GCCACGGGCA CCGGCAAATC    120
CACGCTGGCG TTGCAGCTAG GCCAGCGGCT CGGCGGTGAG ATCGTCAGCG CTGACTCCAT    180
GCAGGTCTAT GAAGGCCTAG ACATCATCAC CAACAAGGTT TCTGCCCAAG AGCAGAGAAT    240
CTGCCGGCAC CACATGATCA GCTTTGTGGA TCCTCTTGTG ACCAATTACA CAGTGGTGGA    300
CTTCAGAAAT AGAGCAACTG CTCTGATTGA AGATATATTT GCCCGAGACA AAATTCCTAT    360
TGTTGTGGGA GGAACCAATT ATTACATTGA ATCTCTGCTC TGGAAAGTTC TTGTCAATAC    420
CAAGCCCCAG GAGATGGGCA CTGAGAAAGT GATTGACCGA AAAGTGGAGC TTGAAAAGGA    480
GGATGGTCTT GTACTTCACA AACGCCTAAG CCAGGTGGAC CCAGAAATGG CTGCCAAGCT    540
GCATCCACAT GACAAACGCA AAGTGGCCAG GAGCTTGCAA GTTTTTGAAG AAACAGGAAT    600
CTCTCATAGT GAATTTCTCC ATCGTCAACA TACGGAAGAA GGTGGTGGTC CCCTTGGAGG    660
TCCTCTGAAG TTCTCTAACC CTTGCATCCT TTGGCTTCAT GCTGACCAGG CAGTTCTAGA    720
TGAGCGCTTG GATAAGAGGG TGGATGACAT GCTTGCTGCT GGGCTCTTGG AGGAACTAAG    780
AGATTTTCAC AGACGCTATA ATCAGAAGAA TGTTTCGGAA AATAGCCAGG ACTATCAACA    840
TGGTATCTTC CAATCAATTG GCTTCAAGGA ATTTCACGAG TACCTGATCA CTGAGGGAAA    900
ATGCACACTG GAGACTAGTA ACCAGCTTCT AAAGAAAGGA CCTGGTCCCA TTGTCCCCCC    960
TGTCTATGGC TTAGAGGTAT CTGATGTCTC GAAGTGGGAG GAGTCTGTTC TTGAACCTGC   1020
TCTTGAAATC GTGCAAAGTT TCATCCAGGG CCACAAGCCT ACAGCCACTC CAATAAAGAT   1080
GCCATACAAT GAAGCTGAGA ACAAGAGAAG TTATCACCTG TGTGACCTCT GTGATCGAAT   1140
CATCATTGGG GATCGCGAAT GGGCAGCGCA CATAAAATCC AAATCCCACT TGAACCAACT   1200
GAAGAAAAGA AGAAGATTGG ACTCAGATGC TGTCAACACC ATAGAAAGTC AGAGTGTTTC   1260
CCCAGACTAT AACAAAGAAC CTAAAGGGAA GGGATCCCCA GGGCAGAATG ATCAAGAGCT   1320
GAAATGCAGC GTTTAAGAGA CATGTCCAGT GGCCTTTGGA AAGGTGGTGG GGATCCAGTT   1380
CAGGAGGGAG GGGTATGTTT GTCTCCCAGT CTGGGCAAAG GAGTGCTATG CGGAATTCTC   1440
TGCATAGCAG AAAAGCTCCC ACCATTTTCT TTTGATGTGG TTTTAAAGTC TCACGTTCTC   1500
TATAATAGAA ACAGCAGGTC TTGTCAGCTC CTTGTGTGGC TGATGTGTCT GGAAATGATG   1560
TAGTTCAGGA AAGCATTTTT TTTTTCTTTG AACCTTAAAG GTTCTATTAT TAAAAGCAGC   1620
ACAGATTCCA CATTTTTATA CATGAGGATC TTCTTTGTGG TGAATACCAG GATTGACTGC   1680
ATCCCTTTAA AAGAAGTTTT ATGTCCCTGA CTCTGGCTAA AATTATCTAA TTTCCAGATG   1740
CTTTTGTAGA TGACTGAAGT ATTTGTGAGC CACATATTGG GAGTTCTAGA TTTGAGTGAA   1800
TGGCAGGAAA GGGCCATCTC CATTGAGATG ATTAAGTGAA CCAAACTAGT TCTCGGAATT   1860
CTACAGAGAA GGAGGGAATC AGACTGAGGA AGCTGTGACA TAGGACTTGA AGACCAAAGA   1920
CTTTGAAATT TGCGAGCTGC TCATGTGTGA GTTATTATCA CTGCTGTCTT TCTATTGAGT   1980
```

TACAAATCTA TATTTTTATT GAAGTTTAAA TAAAGAAAAA ATTTACAAGA AAAAAAAAAA 2040
A 2041

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 892 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```
Met Phe Arg Lys Leu Gly Ser Ser Gly Ser Leu Trp Lys Pro Lys Asn
 1               5                  10                  15
Pro His Ser Leu Glu Tyr Leu Lys Tyr Leu Gln Gly Val Leu Thr Lys
            20                  25                  30
Asn Glu Lys Val Thr Glu Asn Asn Lys Lys Ile Leu Val Glu Ala Leu
        35                  40                  45
Arg Ala Ile Ala Glu Ile Leu Ile Trp Gly Asp Gln Asn Asp Ala Ser
    50                  55                  60
Val Phe Asp Phe Phe Leu Glu Arg Gln Met Leu Leu Tyr Phe Leu Lys
65                  70                  75                  80
Ile Met Glu Gln Gly Asn Thr Pro Leu Asn Val Gln Leu Leu Gln Thr
                85                  90                  95
Leu Asn Ile Leu Phe Glu Asn Ile Arg His Glu Thr Ser Leu Tyr Phe
            100                 105                 110
Leu Leu Ser Asn Asn His Val Asn Ser Ile Ile Ser His Lys Phe Asp
        115                 120                 125
Leu Gln Asn Asp Glu Ile Met Ala Tyr Tyr Ile Ser Phe Leu Lys Thr
    130                 135                 140
Leu Ser Phe Lys Leu Asn Pro Ala Thr Ile His Phe Phe Phe Asn Glu
145                 150                 155                 160
Thr Thr Glu Glu Phe Pro Leu Leu Val Glu Val Leu Lys Leu Tyr Asn
                165                 170                 175
Trp Asn Glu Ser Met Val Arg Ile Ala Val Arg Asn Ile Leu Leu Asn
            180                 185                 190
Ile Val Arg Val Gln Asp Asp Ser Met Ile Ile Phe Ala Ile Lys His
        195                 200                 205
Thr Lys Glu Tyr Leu Ser Glu Leu Ile Asp Ser Leu Val Gly Leu Ser
    210                 215                 220
Leu Glu Met Asp Thr Phe Val Arg Ser Ala Glu Asn Val Leu Ala Asn
225                 230                 235                 240
Arg Glu Arg Leu Arg Gly Lys Val Asp Asp Leu Ile Asp Leu Ile His
                245                 250                 255
Tyr Ile Gly Glu Leu Leu Asp Val Glu Ala Val Ala Glu Ser Leu Ser
            260                 265                 270
Ile Leu Val Thr Thr Arg Tyr Leu Ser Pro Leu Leu Leu Ser Ser Ile
        275                 280                 285
Ser Pro Arg Arg Asp Asn His Ser Leu Leu Leu Thr Pro Ile Ser Ala
    290                 295                 300
Leu Phe Phe Phe Ser Glu Phe Leu Leu Ile Val Arg His His Glu Thr
305                 310                 315                 320
Ile Tyr Thr Phe Leu Ser Ser Phe Leu Phe Asp Thr Gln Asn Thr Leu
                325                 330                 335
```

```
Thr Thr His Trp Ile Arg His Asn Glu Lys Tyr Cys Leu Glu Pro Ile
            340                 345                 350
Thr Leu Ser Ser Pro Thr Gly Glu Tyr Val Asn Glu Asp His Val Phe
        355                 360                 365
Phe Asp Phe Leu Leu Glu Ala Phe Asp Ser Ser Gln Ala Asp Asp Ser
    370                 375                 380
Lys Ala Phe Tyr Gly Leu Met Leu Ile Tyr Ser Met Phe Gln Asn Asn
385                 390                 395                 400
Ala Asp Val Gly Glu Leu Leu Ser Ala Ala Asn Phe Pro Val Leu Lys
                405                 410                 415
Glu Ser Thr Thr Thr Ser Leu Ala Gln Gln Asn Leu Ala Arg Leu Arg
            420                 425                 430
Ile Ala Ser Thr Ser Ser Ile Ser Lys Arg Thr Arg Ala Ile Thr Glu
        435                 440                 445
Ile Gly Val Glu Ala Thr Glu Glu Asp Glu Ile Phe His Asp Val Pro
    450                 455                 460
Glu Glu Gln Thr Leu Glu Asp Leu Val Asp Asp Val Leu Val Asp Thr
465                 470                 475                 480
Glu Asn Ser Ala Ile Ser Asp Pro Glu Pro Lys Asn Val Glu Ser Glu
                485                 490                 495
Ser Arg Ser Arg Phe Gln Ser Ala Val Asp Glu Leu Pro Pro Pro Ser
            500                 505                 510
Thr Ser Gly Cys Asp Gly Arg Leu Phe Asp Ala Leu Ser Ser Ile Ile
        515                 520                 525
Lys Ala Val Gly Thr Asp Asp Asn Arg Ile Arg Pro Ile Thr Leu Glu
    530                 535                 540
Leu Ala Cys Leu Val Ile Arg Gln Ile Leu Met Thr Val Asp Asp Glu
545                 550                 555                 560
Lys Val His Thr Ser Leu Thr Lys Leu Cys Phe Glu Val Arg Leu Lys
                565                 570                 575
Leu Leu Ser Ser Ile Gly Gln Tyr Val Asn Gly Glu Asn Leu Phe Leu
            580                 585                 590
Glu Trp Phe Glu Asp Glu Tyr Ala Glu Phe Glu Val Asn His Val Asn
        595                 600                 605
Phe Asp Ile Ile Gly His Glu Met Leu Leu Pro Pro Ala Ala Thr Pro
    610                 615                 620
Leu Ser Asn Leu Leu Leu His Lys Arg Leu Pro Ser Gly Phe Glu Glu
625                 630                 635                 640
Arg Ile Arg Thr Gln Ile Val Phe Tyr Leu His Ile Arg Lys Leu Glu
                645                 650                 655
Arg Asp Leu Thr Gly Glu Gly Asp Thr Glu Leu Pro Val Arg Val Leu
            660                 665                 670
Asn Ser Asp Gln Glu Pro Val Ala Ile Gly Asp Cys Ile Asn Leu His
        675                 680                 685
Asn Ser Asp Leu Leu Ser Cys Thr Val Val Pro Gln Gln Leu Cys Ser
    690                 695                 700
Leu Gly Lys Pro Gly Asp Arg Leu Ala Arg Phe Leu Val Thr Asp Arg
705                 710                 715                 720
Leu Gln Leu Ile Leu Val Glu Pro Asp Ser Arg Lys Ala Gly Trp Ala
                725                 730                 735
Ile Val Arg Phe Val Gly Leu Leu Gln Asp Thr Thr Ile Asn Gly Asp
            740                 745                 750
Ser Thr Asp Ser Lys Val Leu His Val Val Val Glu Gly Gln Pro Ser
        755                 760                 765
Arg Ile Lys Lys Arg His Pro Val Leu Thr Ala Lys Phe Ile Phe Asp
    770                 775                 780
Asp His Ile Arg Cys Met Ala Ala Lys Gln Arg Leu Thr Lys Gly Arg
785                 790                 795                 800
```

```
Gln Thr Ala Arg Gly Leu Lys Leu Gln Ala Ile Cys Ser Ala Leu Gly
                805                 810                 815
Val Pro Arg Ile Asp Pro Ala Thr Met Thr Ser Ser Pro Arg Met Asn
            820                 825                 830
Pro Phe Arg Ile Val Lys Gly Cys Ala Pro Gly Ser Val Arg Lys Thr
        835                 840                 845
Val Ser Thr Ser Ser Ser Ser Ser Gln Gly Arg Pro Gly His Tyr Ser
    850                 855                 860
Ala Asn Leu Arg Ser Ala Ser Arg Asn Ala Gly Met Ile Pro Asp Asp
865                 870                 875                 880
Pro Thr Gln Pro Ser Ser Ser Ser Glu Arg Arg Ser
                885                 890
```

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 355 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```
Met Ala Glu Lys Ala Glu Asn Leu Pro Ser Ser Ser Ala Glu Ala Ser
 1                5                  10                  15
Glu Glu Pro Ser Pro Gln Thr Gly Pro Asn Val Asn Gln Lys Pro Ser
            20                  25                  30
Ile Leu Val Leu Gly Met Ala Gly Ser Gly Lys Thr Thr Phe Val Gln
        35                  40                  45
Arg Leu Thr Ala Phe Leu His Ala Arg Lys Thr Pro Pro Tyr Val Ile
    50                  55                  60
Asn Leu Asp Pro Ala Val Ser Lys Val Pro Tyr Pro Val Asn Val Asp
65                  70                  75                  80
Ile Arg Asp Thr Val Lys Tyr Lys Glu Val Met Lys Glu Phe Gly Met
                85                  90                  95
Gly Pro Asn Gly Ala Ile Met Thr Cys Leu Asn Leu Met Cys Thr Arg
            100                 105                 110
Phe Asp Lys Val Ile Glu Leu Ile Asn Lys Arg Ser Ser Asp Phe Ser
        115                 120                 125
Val Cys Leu Leu Asp Thr Pro Gly Gln Ile Glu Ala Phe Thr Trp Ser
    130                 135                 140
Ala Ser Gly Ser Ile Ile Thr Asp Ser Leu Ala Ser Ser His Pro Thr
145                 150                 155                 160
Val Val Met Tyr Ile Val Asp Ser Ala Arg Ala Thr Asn Pro Thr Thr
                165                 170                 175
Phe Met Ser Asn Met Leu Tyr Ala Cys Ser Ile Leu Tyr Arg Thr Lys
            180                 185                 190
Leu Pro Phe Ile Val Val Phe Asn Lys Ala Asp Ile Val Lys Pro Thr
        195                 200                 205
Phe Ala Leu Lys Trp Met Gln Asp Phe Glu Arg Phe Asp Glu Ala Leu
    210                 215                 220
Glu Asp Ala Arg Ser Ser Tyr Met Asn Asp Leu Ser Arg Ser Leu Ser
225                 230                 235                 240
Leu Val Leu Asp Glu Phe Tyr Cys Gly Leu Lys Thr Val Cys Val Ser
                245                 250                 255
```

```
Ser Ala Thr Gly Glu Gly Phe Glu Asp Val Met Thr Ala Ile Asp Glu
            260                 265                 270
Ser Val Glu Ala Tyr Lys Lys Glu Tyr Val Pro Met Tyr Glu Lys Val
        275                 280                 285
Leu Ala Glu Lys Lys Leu Leu Asp Glu Glu Glu Arg Lys Lys Arg Asp
    290                 295                 300
Glu Glu Thr Leu Lys Gly Lys Ala Val His Asp Leu Asn Lys Val Ala
305                 310                 315                 320
Asn Pro Asp Glu Phe Leu Glu Ser Glu Leu Asn Ser Lys Ile Asp Arg
                325                 330                 335
Ile His Leu Gly Gly Val Asp Glu Glu Asn Glu Glu Asp Ala Glu Leu
            340                 345                 350
Glu Arg Ser
        355
```

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 434 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

```
Met Ser Glu Lys Thr Phe His Lys Ala Gln Thr Ile Arg Ala Lys Ala
 1               5                  10                  15
Ser Gly Val Pro Ser Ile Val Glu Ala Val Gln Phe His Gly Val Arg
            20                  25                  30
Ile Thr Lys Asn Asp Ala Leu Val Lys Glu Val Ser Glu Leu Tyr Arg
        35                  40                  45
Ser Lys Asn Leu Asp Glu Leu Val His Asn Ser His Leu Ala Ala Arg
    50                  55                  60
His Leu Gln Glu Val Gly Leu Met Asp Asn Ala Val Ala Leu Ile Asp
65                  70                  75                  80
Thr Ser Pro Ser Ser Asn Glu Gly Tyr Val Val Asn Phe Leu Val Arg
                85                  90                  95
Glu Pro Lys Ser Phe Thr Ala Gly Val Lys Ala Gly Val Ser Thr Asn
            100                 105                 110
Gly Asp Ala Asp Val Ser Leu Asn Ala Gly Lys Gln Ser Val Gly Gly
        115                 120                 125
Arg Gly Glu Ala Ile Asn Thr Gln Tyr Thr Tyr Thr Val Lys Gly Asp
    130                 135                 140
His Cys Phe Asn Ile Ser Ala Ile Lys Pro Phe Leu Gly Trp Gln Lys
145                 150                 155                 160
Tyr Ser Asn Val Ser Ala Thr Leu Tyr Arg Ser Leu Ala His Met Pro
                165                 170                 175
Trp Asn Gln Ser Asp Val Asp Glu Asn Ala Ala Val Leu Ala Tyr Asn
            180                 185                 190
Gly Gln Leu Trp Asn Gln Lys Leu Leu His Gln Val Lys Leu Asn Ala
        195                 200                 205
Ile Trp Arg Thr Leu Arg Ala Thr Arg Asp Ala Ala Phe Ser Val Arg
    210                 215                 220
Glu Gln Ala Gly His Thr Leu Lys Phe Ser Leu Glu Asn Ala Val Ala
225                 230                 235                 240
```

```
Val Asp Thr Arg Asp Arg Pro Ile Leu Ala Ser Arg Gly Ile Leu Ala
                245                 250                 255
Arg Phe Ala Gln Glu Tyr Ala Gly Val Phe Gly Asp Ala Ser Phe Val
            260                 265                 270
Lys Asn Thr Leu Asp Leu Gln Ala Ala Ala Pro Leu Pro Leu Gly Phe
        275                 280                 285
Ile Leu Ala Ala Ser Phe Gln Ala Lys His Leu Lys Gly Leu Gly Asp
    290                 295                 300
Arg Glu Val His Ile Leu Asp Arg Cys Tyr Leu Gly Gly Gln Gln Asp
305                 310                 315                 320
Val Arg Gly Phe Gly Leu Asn Thr Ile Gly Val Lys Ala Asp Asn Ser
                325                 330                 335
Cys Leu Gly Gly Gly Ala Ser Leu Ala Gly Val Val His Leu Tyr Arg
            340                 345                 350
Pro Leu Ile Pro Pro Asn Met Leu Phe Ala His Ala Phe Leu Ala Ser
        355                 360                 365
Gly Ser Val Ala Ser Val His Ser Lys Asn Leu Val Gln Gln Leu Gln
    370                 375                 380
Asp Thr Gln Arg Val Ser Ala Gly Phe Gly Leu Ala Phe Val Phe Lys
385                 390                 395                 400
Ser Ile Phe Arg Leu Glu Leu Asn Tyr Thr Tyr Pro Leu Lys Tyr Val
                405                 410                 415
Leu Gly Asp Ser Leu Leu Gly Gly Phe His Ile Gly Ala Gly Val Asn
            420                 425                 430
Phe Leu
```

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 198 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

```
Met Leu Tyr Ile Leu Trp Lys Leu Asn Tyr Leu Gln Lys Lys Met Ser
 1               5                  10                  15
Leu Arg Lys Ile Asn Phe Val Thr Gly Asn Val Lys Lys Leu Glu Glu
            20                  25                  30
Val Lys Ala Ile Leu Lys Asn Phe Glu Val Ser Asn Val Asp Val Asp
        35                  40                  45
Leu Asp Glu Phe Gln Gly Glu Pro Glu Phe Ile Ala Glu Arg Lys Cys
    50                  55                  60
Arg Glu Ala Val Glu Ala Val Lys Gly Pro Val Leu Val Glu Asp Thr
65                  70                  75                  80
Ser Leu Cys Phe Asn Ala Met Gly Gly Leu Pro Gly Pro Tyr Ile Lys
                85                  90                  95
Trp Phe Leu Lys Asn Leu Lys Pro Glu Gly Leu His Asn Met Leu Ala
            100                 105                 110
Gly Phe Ser Asp Lys Thr Ala Tyr Ala Gln Cys Ile Phe Ala Tyr Thr
        115                 120                 125
Glu Gly Leu Gly Lys Pro Ile His Val Phe Ala Gly Lys Cys Pro Gly
    130                 135                 140
Gln Ile Val Ala Pro Arg Gly Asp Thr Ala Phe Gly Trp Asp Pro Cys
145                 150                 155                 160
```

Phe Gln Pro Asp Gly Phe Lys Glu Thr Phe Gly Glu Met Asp Lys Asp
165 170 175
Val Lys Asn Glu Ile Ser His Arg Ala Lys Ala Leu Glu Leu Leu Lys
180 185 190
Glu Tyr Phe Gln Asn Asn
195

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

CGAACACTTT ATATTTCTCG 20

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GATAGTTCCC TTCGTTCGGG 20

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

TTTCTGGATT TTAACCTTCC 20

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

TTTCCGAGAA GTCACGTTGG 20

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

TACAGGAATT TTTGAACGGG 20

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

CTTCAGATGA CGTGGATTCC 20

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

GGAATCCGAA AAAGTGAACT 20

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

AAGAGATACA CTCAATGGGG 20

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

ATCGATACCA CCGTCTCTGG 20

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

TTGAATCTAC ACTAATCACC 20

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

CCAATTATCT TTTCCAGTCA 20

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

ACATTATAAA GTTACTGTCC 20

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

TTTTAGTTAA AGCATTGACC 20

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

ACATCTTTAT CCATTTCTCC 20

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

TGCAAAGGCT CTGGAACTCC 20

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

AAAAACCACT TGATATAAGG 20

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

CATCCAAAAG CAGTATCACC 20

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

TTAATTGGAT GCAAGCACCC C 21

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

ATTACTATAC GAACATTTCC 20

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

TTGTAAAGGC GTTAGTTTGG 20

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

CAGGAGTATT TGGTGATGCG 20

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

CGACGGGGAG AAGGTGACGG 20

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

AAAACTTCTA CCAACAATGG 20

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

CGTAATCTCT CTCGATTAGC 20

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

CCGTGGGATG GCTACTTGCC 20

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

TGGATTTGTG GCACGAGCGG 20

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

TTGATTGCCT CTCCTCGTCC 20

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

ATCAACATCT GATTGATTCC 20

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

CAGCGAGCGC ATGCAACTAT ATATTGAGCA GG 32

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 41 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

AATAAATATT TAAATATTCA GATATACCCT GAACTCTACA G 41

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 45 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

AAACTGTAGA GTTCAGGGTA TATCTGAATA TTTAAATATT TATTC 45

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 34 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

GTACGTGGAG CTCTGCAACT ATATATTGAG CAGG 34

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

ATGACACTGC AGGATAGTTC CCTTCGTTCG GG 32

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

GTGTTGCATC AGTTCATTCC 20

(2) INFORMATION FOR SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

GCTGTGCTAG AAGTCAGAGG 20

(2) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

GTTCTCCTTG GAATTCATCC 20

(2) INFORMATION FOR SEQ ID NO:44:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

AGTATATCTA GATGTGCGAG TCTCTGCCAA TT 32

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

AGTAATTGTA CATTTAGTGG 20

(2) INFORMATION FOR SEQ ID NO:46:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

ATTAACCTTA CTTACTTACC 20

(2) INFORMATION FOR SEQ ID NO:47:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

CTAAACTAAG TAATATAACC 20

(2) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

GTTGATTCTT TGAGCACTGG 20

(2) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

AATTCGACCA ATTACATTGG 20

(2) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

AACATAGTTG TTGAGGAAGG 20

(2) INFORMATION FOR SEQ ID NO:51:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

AATTAATGGA GATTCTACGG 20

(2) INFORMATION FOR SEQ ID NO:52:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

TCAGCATCTA GAAATGCAGG 20

(2) INFORMATION FOR SEQ ID NO:53:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

CGAATGTCAA CATTCACTGG 20

(2) INFORMATION FOR SEQ ID NO:54:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

CTTAACCTGA TGTGTACTCG 20

(2) INFORMATION FOR SEQ ID NO:55:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

ATGAAGCTTT AGAGGATGCC 20

(2) INFORMATION FOR SEQ ID NO:56:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

CGACGAATTT CTGGAGTCGG 20

(2) INFORMATION FOR SEQ ID NO:57:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

ACTGCATTAT CCATTAATCC 20

(2) INFORMATION FOR SEQ ID NO:58:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

CACCCAAATA ACATCTATCC 20

(2) INFORMATION FOR SEQ ID NO:59:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

TTTAACCTCA TCTTCGCTGG 20

(2) INFORMATION FOR SEQ ID NO:60:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

ATGTTCCGCA AGCTTGGTTC 20

(2) INFORMATION FOR SEQ ID NO:61:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

TTTAATTACC CAAGTTTGAG 20

(2) INFORMATION FOR SEQ ID NO:62:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

TTTTAACCCA GTTACTCAAG 20

INTERNATIONAL SEARCH REPORT

Int ional Application No
PCT/CA 98/00803

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N9/10 C12Q1/68 A01K67/027 //C12N15/62

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12N C12Q A01K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WILSON R ET AL: "2.2 MB OF CONTIGUOUS NUCLEOTIDE SEQUENCE FROM CHROMOSOME III OF C ELEGANS" NATURE, vol. 368, no. 6466, 3 March 1994, pages 32-38, XP002029739	1-7,9, 11-15
Y	see the whole document -& DATABASE EMBL - CEZC395 Entry CEZC395, Acc.No. U13642, 30 November 1994 WILSON, R. ET EL.: "Caenorhabditis elegans cosmid ZC395" XP002089006 see the whole document -& DATABASE EMBL - EMINV Entry CEC34E10, Acc.No. U10402, 30 June 1994 WILSON, R. ET EL.: "Caenorhabditis elegans cosmid C34E10"	8

-/--

[X] Further documents are listed in the continuation of box C. [ ] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
11 January 1999	22/01/1999

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Smalt, R

3

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Inte onal Application No
PCT/CA 98/00803

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
	XP002089545 see the whole document ---	
Y	ADAMS M D ET AL: "INITIAL ASSESSMENT OF HUMAN GENE DIVERSITY AND EXPRESSION PATTERNS BASED UPON 83 MILLION NUCLEOTIDES OF CDNA SEQUENCE" NATURE, vol. 377, 28 September 1995, pages 3-17, XP002042918 see the whole document -& DATABASE EMBL - EMEST14 Entry HSZZ37212, Acc.No. AA332152, 18 April 1997 ADAMS, M.D. ET AL.: "EST36068 Embryo, 8 week I Homo sapiens cDNA 5' end similar to similar to tRNA isopentenyltransferase." XP002089546 see the whole document -& DATABASE EMBL - EMEST14 Entry HSZZ61218, Acc.No. AA356092, 18 April 1997 ADAMS, M.D. ET AL.: "EST64588 Jurkat T-cells VI Homo sapiens cDNA 5' end similar to similar to tRNA isopentenyltransferase." XP002089547 see the whole document ---	8
A	LAKOWSKI, B. ET AL.: "Determination of life-span in Caenorhabditis elegans by four clock genes." SCIENCE, vol. 272, 17 May 1996, pages 1010-3, XP002089004 cited in the application see the whole document ---	
A	EWBANK, J.J. ET AL.: "Structural and functional conservation of the Caenorhabditis elegans timing gene clk-1." SCIENCE, vol. 275, 14 February 1997, pages 980-3, XP002089005 cited in the application see the whole document ---	
A	SPIETH, J. ET AL.: "Operon in C. elegans: polycistronic mRNA precursors are processed by trans-splicing of SL2 to downstream coding regions." CELL, vol. 73, 1993, pages 521-32, XP002089544 cited in the application see the whole document -----	

3

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CA 98/ 00803

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [X] Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
SEE FURTHER INFORMATION SHEET PCT/ISA/210

2. [ ] Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. [ ] Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. [ ] As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. [ ] As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. [ ] No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

[ ] The additional search fees were accompanied by the applicant's protest.

[ ] No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Although claims 18-27 are directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.

The claims 18-27, referring to compounds interfering with the enzymatic activity of the claimed proteins, could not be searched completely due to the lack of support of these compounds in the application.